NAVAL POSTGRADUATE SCHOOL

Monterey, California

DISSERTATION

An Investigation of Nonlinear Controls and Regression-Adjusted Estimators for Variance Reduction in Computer Simulation

by

Richard L. R[illegible]r

March, 1991

Dissertation Advisor: Peter A.W. Lewis

REPORT DOCUMENTATION PAGE

1a. Report Security Classification: Unclassified	1b. Restrictive Markings
2a. Security Classification Authority	3. Distribution/Availability of Report: Approved for public release; distribution is unlimited.
2b. Declassification/Downgrading Schedule	
4. Performing Organization Report Number(s)	5. Monitoring Organization Report Number(s)
6a. Name of Performing Organization: Naval Postgraduate School; 6b. Office Symbol: OR	7a. Name of Monitoring Organization: Naval Postgraduate School
6c. Address (*City, State, and ZIP code*): Monterey, CA 93943-5000	7b. Address (*City, State, and ZIP code*): Monterey, CA 93943-5000
8a. Name of Funding/Sponsoring Organization; 8b. Office Symbol	9. Procurement Instrument Identification Number
8c. Address (*City, State, and ZIP code*)	10. Source of Funding Numbers: Program Element No; Project No; Task No; Work Unit Accession No

11. Title (*Include Security Classification*)
AN INVESTIGATION OF NONLINEAR CONTROLS AND REGRESSION-ADJUSTED ESTIMATORS FOR VARIANCE REDUCTION IN COMPUTER SIMULATION

12. Personal Author(s): Richard L. Ressler

13a. Type of Report	13b. Time covered	14. Date of Report (*year, month, day*)	15. Page Count
Ph.D. Dissertation	From To	March 1991	172

16. Supplementary Notation: The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

17. Cosati Codes			18. Subject Terms (*continue on reverse if necessary and identify by block number*)
Field	Group	Subgroup	Variance Reduction, Nonlinear Controls, Quantiles, Queueing Simulation, Regenerative Simulation, Asymptotic Expansions

19. Abstract (*continue on reverse if necessary and identify by block number*)

This dissertation develops new techniques for variance reduction in computer simulation. It demonstrates that applying nonlinear transformations to control variables can increase their effectiveness over linear controls. It shows how one can reduce the variance of quantile estimates, where the quantile of interest is a continuous and strictly monotone transformation of the control quantile, by transforming the control quantile with a different continuous and strictly monotone transformation. Asymptotic expansions are developed to validate the improved performance of the nonlinear control for the quantile estimate. Finally, in the realm of regenerative simulation, regression-adjusted techniques are applied to controlled regenerative estimates. The resulting estimates have a greatly reduced estimated mean square error.

20. Distribution/Availability of Abstract	21. Abstract Security Classification	
☒ unclassified/unlimited ☐ same as report ☐ DTIC users	Unclassified	
22a. Name of Responsible Individual: Peter A. W. Lewis	22b. Telephone (*include Area code*): (408) 646-2283	22c. Office Symbol: ORLw

DD FORM 1473, 84 MAR — Previous editions are obsolete. — Security Classification of this Page: Unclassified

Approved for public release; distribution is unlimited

An Investigation of Nonlinear Controls and Regression-Adjusted Estimators for Variance Reduction in Computer Simulation

by

Richard L. Ressler

Major, United States Army

B. A., University of Pennsylvania, 1978

M.S., Naval Postgraduate School, 1987

Submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL

March, 1991

Author: _____

Richard L. Ressler

Approved by: _____

Alan R. Washburn
Professor of Operations Research

_____

R. Kevin Wood
Associate Professor of Operations Research

_____

Maurice Weir
Professor of Mathematics

_____

Jerome H. Friedman
Professsor of Statistics
Stanford University

_____

Peter A.W. Lewis
Professor of Operations Research
Dissertation Supervisor

Approved by: _____

Peter Purdue, Chairman, Department of Operations Research

Approved by: _____

Richard S. Elster, Dean of Instruction

ABSTRACT

This dissertation develops new techniques for variance reduction in computer simulation. It demonstrates that applying nonlinear transformations to control variables can increase their effectiveness over linear controls. It shows how one can reduce the variance of quantile estimates, where the quantile of interest is a continuous and strictly monotone transformation of the control quantile, by transforming the control quantile with a different continuous and strictly monotone transformation. Asymptotic expansions are developed to validate the improved performance of the nonlinear control for the quantile estimate. Finally, in the realm of regenerative simulation, regression-adjusted techniques are applied to controlled regenerative estimates. The resulting estimates have a greatly reduced estimated mean square error.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENT

I thank my wife, Marianne, and my children, Christopher and Alexandra.

I. INTRODUCTION

This dissertation is an investigation of the use of new techniques such as data transformations, the Alternating Conditional Expectation (ACE) algorithm of Breiman and Friedman (1985), and regression-adjusted estimates to make nonlinear controls a useful method for reducing the variance of estimates (of unknown constants) produced via Monte Carlo computer simulations i.e., simulations which use pseudo-random numbers. People use variance reduction techniques to either save computer resources or improve the precision of the estimates gained for the same amount of resources. The following sections provide a short background on Monte Carlo computer simulation, discuss the new variance reduction techniques described in the dissertation and briefly outline the chapters that follow.

A. BACKGROUND

While there are many definitions of types of "simulations" (see Kleijnen, 1974, Chap. 1), the emphasis in this work is on extending variance reduction techniques for stochastic simulations. One can recognize these types of simulations by the fact that a pseudo-random number generator provides input essential to the simulation so that both the inputs to the simulation and the output from the simulation are random variables. The stochastic simulations in mind do not include "physical simulations" such as flight simulators, or "deterministic simulations" such as using a computer to solve large sets of partial differential equations over time as is done to study chemical interactions. The two types of stochastic simulations in mind have been called "statistical" simulations and "discrete-event system simulations."

Statistical simulations are concerned with deriving information about the "behavior of statistical estimates or procedures as a function of sample size, population distribution and other factors," while system simulations are concerned with deriving information about how a process, perhaps physical, responds over a range of input parameters (Lewis and Orav, 1989, Chap. 8). These simulations can both be considered as "controlled statistical sampling techniques" (Lewis and Orav, 1989, p. 9) in that one uses a computer to

simulate a given random process using pseudo-random numbers in order to provide samples (replications) of information of interest about the process. The two types of simulations differ in scale in that most (but definitely not all) statistical simulations could be solved on a large computer in reasonable amounts of time so that variance reduction techniques are not needed and are an inefficient use of the analyst's time. System simulations, on the other hand, especially when there are many factors influencing the output that need to be examined, can use all the help they can get, even if it is, for example, a modest one-third reduction in computing time for a given precision of output.

An important point is that one must develop the computer simulation so that using the numerical output from the simulation, one can calculate a statistic that estimates the quantity of interest. As a random variable, this estimate has its own associated probability distribution with its own mean and variance. To know how "good" this estimate is, or to conduct meaningful comparisons using the estimate, *it is absolutely essential that one have an estimate of the precision of the estimate.* This is an extension, in complex simulation, of the simple idea that in using the sample mean, $\bar{x}$, of an i.i.d. sample $x_1, \ldots, x_n$, to estimates $\mathrm{E}[X]$, it is known that the $\mathrm{Var}[\bar{x}] = \sigma_x^2/n$, and moreover, this can be estimated by s_x^2/n where s_x^2 is the sample variance.

A common measure of the precision of an estimate is a relative measure; the standard deviation of the point estimate divided by the absolute value of the point estimate (Lewis and Orav, 1989, p. 10) provided that the absolute value is not zero. When discussing variance reduction, one is concerned about comparing the standard deviation or variance of estimates of the same quantity. Thus in what follows, the standard deviation or variance will be considered as absolute measures of precision; the common value of the underlying quantity factors out.

One can often summarize the information of interest desired from the simulation via a numerical quantification. These quantifications may include the mean and variance of a random variable or the parameter of a probability distribution. Quantiles are another class of quantifications useful in particular for characterizing the extremes of a distribution. The α-quantile of a probability distribution for a random variable Y is the smallest number y_α such that the probability that $Y \leq y$ is greater than or equal to α. Quantiles of interest may include the .5 quantile (median), as a general measure of location, or the

extreme quantiles such as the .95 or .99 quantile. Estimation of quantiles is fundamentally different from estimating means in that the order-statistic-based quantile estimator is a nondifferentiable function of the data and the estimate is not representable as some kind of sample mean. Thus, in this dissertation, different techniques are investigated for applying nonlinear controls both to quantile estimates and to other quantifications.

Regardless of whether one uses the relative or the absolute measure of precision for the estimate, the method of calculating the estimate must provide in addition to the point estimate of the quantity of interest (the statistic itself), a reliable estimate of the standard deviation (variance) of the point estimate, such as s. Without this estimate of the precision, the point estimate is virtually useless. In general, the smaller the variance of the estimate the better. The most basic way of reducing the variance of the estimate is to increase the sample size i.e., the number of replications of the simulation.

For many statistics, such as the sample mean, the variance of the statistic decreases linearly as a function of the sample size n. Thus one can reduce the variance of the estimate by simply increasing n, the sample size of the simulation. Unfortunately, the standard deviation of the estimate, which is measured on the same scale as the quantification of interest, decreases at a rate of $1/\sqrt{n}$; thus to reduce the standard deviation of the estimate by a factor of one half, one must quadruple the sample size. For many computer simulations, such as a complex combat model, having to increase the precision to a reasonable level may create the predicament of expending an inordinate amount of resources. Even fairly simple simulations such as an M/M/1 queue may require a huge amount of resources to quadruple the sample size when the traffic intensity is above .99. This predicament of having to consume large amounts of resources in order to increase precision is a major motivation for the development and use of variance reduction techniques.

One can often use so called "variance reduction techniques" to attempt to reduce the standard deviation of the estimate with only a moderate investment of resources. There is a wide variety of variance reduction techniques available (see Lewis and Orav, 1989, Chap. 11 or Kleijnen, 1974, Chap. III). In general, variance reduction techniques require the investment of resources in additional analytical work and more complex programming for the computer simulation. This investment of resources may yield the benefit of a reduced

estimate of the standard deviation for the same sample size, or if desired, achieving the same precision using a smaller sample size than otherwise required.

Linear controls are one of several variance reduction techniques and use known information about other random variables present in the simulation to reduce the variance of the estimate. To use linear controls, one must know the expected value of a random variable (the control) that is correlated with the statistic of interest. The method exploits the knowledge of the expected value and the correlation to reduce the variance. Unfortunately, the effectiveness of a linear control scheme depends upon the amount of correlation between the control variable and the statistic of interest. When the relationship between the statistic of interest and the control is not very linear, so that the correlation is low, a linear control may have little effect in reducing the variance. Although Lavenberg and Welch (1981) cite several papers with illustrative examples of the use of linear controls in their survey paper, they "were unable to find any published report describing the application of control variables in an actual practical environment." One possible reason for this lack of use may be that many potential controls are relatively ineffective because they have a nonlinear relationship with the statistic of interest which is not captured by a linear control.

One potential remedy for the lack of effectiveness is to use nonlinear transformations and create nonlinear controls for variance reduction. Breiman and Friedman (1985) showed that transformations exist that will maximize the squared correlation between one transformed response variable (the statistic of interest) and a linear combination of transformed predictor variables (a set of controls). Breiman and Friedman (1985) also developed the Alternating Conditional Expectation (ACE) algorithm for nonparametrically estimating these optimal transformations from a set of data. ACE also provides an estimate of the maximum correlation one can obtain between the transformed response and the combination of transformed predictors. Consequentially it can be used to given an upper bound on the achievable variance reduction for a given set of controls, as well as to suggest a nonlinear transformation of the control that will maximize the correlation one can obtain.

Another potential remedy for insufficient precision in simulation involves the use of asymptotic expansions for the bias of an estimator. One can use linearly or nonlinearly controlled estimates to compute the coefficients in the expansions and produce an estimator

with reduced bias and variance. This idea is developed here in the framework of a special type of system simulation known as "regenerative simulation." The result of using the controlled estimates together with the asymptotic expansions is the "average regression-adjusted controlled regenerative estimate."

B. CONTRIBUTIONS OF THIS DISSERTATION

The potential remedies of nonlinear controls and regression-adjusted controlled regenerative estimates are new contributions to the realm of variance reduction techniques. Linear control schemes have included nonlinear transformations where the parameters were fixed i.e., the use of C^2 in place of or as well as C as a control. The use of nonlinear transformations of control variables where the nonlinear parameters are unspecified *a priori* and determined from the data is a new and truly nonlinear approach to variance reduction.

It is shown how one can combine indicator functions using "cutpoints" with transformations of linear controls to create new, more flexible, nonlinear controls. It is also shown that, similar to linear regression, one can use nonlinear least-squares regression to estimate coefficients for the differentiable nonlinear transformations. While nonlinear controls offer the potential of increased effectiveness in reducing variance in simulation, they have a drawback over the linear control of increased difficulty in determining the expected value of the transformed control.

ACE (Breiman and Friedman, 1985) was developed as a tool for nonparametric modelling. A new application of ACE given in this dissertation is in selecting potential controls and transformations in simulations. One can use the ACE estimate of the maximum squared correlation between a potential control and a statistic of i: ·st as a performance standard for the control i.e., if ACE estimates a maximum squared correlation of only .05, the nonlinear control will be ineffective and not worth the effort. ACE can also be used to help select transformations, or in the case of the indicator function transformations, an effective cutpoint.

Several techniques are developed for controlling quantile estimates in simulation. The nonlinear controls for quantiles exploit two properties of quantile estimators. The first is the behavior of quantiles under strictly monotone transformations of the underlying random variable and the second is the asymptotic properties of an order-statistic-based quantile

estimator. These two properties are combined to greatly simplify the burden of computing the expected value of the transformed nonlinear control. The method of sectioning (see Lewis and Orav, 1989, Chap. 9) is chosen from several competitors for estimating the variance of the controlled quantile estimates and it is shown how one can graphically choose the sectioning parameters so as to maximize the effectiveness of the nonlinear control.

An interesting point that arises when controlling quantile estimates concerns the issue of normality. Typically, users of linear controls try to induce a multivariate normal distribution between the statistic of interest and the controls since a linear control is optimal for a multivariate normal distribution (see Nelson, 1988, and Lancaster, 1966) and because normal theory statistical methodology can be used for the simulation output analysis. Since the quantile estimator used here is asymptotically normally distributed, this would involve using large sample sizes.

It is demonstrated here that by selecting the sectioning parameters appropriately, a nonlinear control at a small sample size can produce a more precise (smaller variance) estimate than a linear control at a large sample size, and a stable estimate of the variance. One wants to avoid the asymptotic normality because the effectiveness of the nonlinear control reduces then to that of the linear control. In fact some authors (Glynn and Whitt, 1989) have dismissed nonlinear controls for this reason; it is the thesis of this dissertation that non-normality is acceptable if it leads to significant variance reduction. Moreover, the use of graphical methods in output analysis will lead to more sensible (non-normal) output analysis than a blind reliance on asymptotic normality.

Asymptotic expansions are developed, building on work by David and Johnson (1956), for the variance of monotonically transformed quantile estimators and the squared correlation between two transformed quantile estimators. These expansions are then used to show that with the proper selection of a transformation, a nonlinear control will be more effective than a linear control in reducing the variance of the quantile estimate for finite sample sizes.

The technique of regression-adjusted controlled regenerative estimates results from combining methods presented in two papers. Iglehart and Lewis (1979) developed linear controls for regenerative estimates while Heidleberger and Lewis (1981) developed regression-adjusted regenerative estimates. Regression-adjusted estimates work with estimators that are nonlinear functions of the data such as ratio estimators. Estimates of the stationary

waiting time from an M/M/1 and M/G/1 queue at high traffic intensities are used to show how one can combine the two techniques to compute average regression-adjusted controlled regenerative estimates along with an associated estimate of their variance. The average regression-adjusted controlled regenerative estimates tend to have smaller estimated mean square error than either controlled regenerative estimates or average regression-adjusted regenerative estimates.

C. OUTLINE OF THE DISSERTATION

Chapter II provides a more detailed background on linear controls. The problems of low effectiveness are discussed with the possible solution being the use of nonlinear transformations. A brief description of Breiman and Friedman's (1985) ACE algorithm then follows. Nonlinear controls are defined and discussed next and the sample mean of the Anderson-Darling Statistic (Anderson and Darling, 1952) is used as an example to demonstrate their use. Chapter III discusses several methods of quantile estimation and the use of nonlinear controls for reducing the estimate of the standard deviation for order-statistic-based quantile estimates. Chapter IV develops new asymptotic expansions for the squared correlation between monotonically transformed quantile estimators. An example demonstrates that in certain situations, properly chosen nonlinear controls will decrease the estimate of the standard deviation of the quantile estimate over linear controls. Chapter V uses the framework of the regenerative simulation of queues to demonstrate how one can use nonlinear controls effectively for low traffic intensity queues. Chapter V also details how one can use the regression-adjusted controlled estimators for decreasing the estimated mean square error of an estimate.

D. SUMMARY

Despite improvements in computing speed, computer simulations will continue to grow in complexity and require more resources. One means for reducing the resource requirements is through the use of variance reduction techniques. While the method of linear controls is well known, the effectiveness of linear controls is often limited by low correlation between the control and the statistic of interest. Thus researchers will continue to seek improved, more effective, techniques for variance reduction.

Two methods for improving upon linear controls for variance reduction in simulation are presented in this dissertation. Using nonlinear transformations for nonlinear controls can improve the precision of estimates of both means and quantile from simulations over crude or linearly controlled estimates. Regression-adjusted controlled regenerative estimates can improve the precision and bias in regenerative estimates more than using controls or regression-adjusted techniques alone. These techniques provide researchers with potential tools for analyzing their simulation results.

II. LINEAR AND NONLINEAR CONTROLS

A. INTRODUCTION

This chapter develops the use of nonlinear control variables for reducing the variance of an estimate of a mean from a Monte Carlo computer simulation. The first section discusses the variance reduction technique of linear controls and why using nonlinear transformations may improve their effectiveness for variance reduction. The second section defines nonlinear controls and explains the usefulness of Breiman and Friedman's (1985) ACE algorithm for estimating optimal nonlinear transformations and bounding the gains in variance reduction that might be achieved with controls. The next section details some possible methods for introducing nonlinearity into a control function in order to approximate the optimal nonlinear transformation. The following section shows, using the sample analog to the variance reduction formula, that nonlinear least-squares regression may be used to find the parameters of the transformations. Finally, the last section provides an example of the use of nonlinear controls to reduce the variance of an estimate of a mean.

B. LINEAR CONTROLS

This section provides a short introduction to the well-known variance reduction technique of linear controls. Many books on simulation contain information on linear controls. One can find further details on linear controls in Lewis and Orav (1989). Assume that one is using a Monte Carlo computer simulation to estimate an unknown quantity. Let Y be the statistic of interest that is calculated from the simulation output to estimate $\mathrm{E}[Y]$. Assume there exists (by design of the simulation) a vector $\underline{C}$ that consists of $q \geq 1$ random variables C_j, for $j = 1, \ldots, q$. Also assume that the C_j are correlated with (related to or associated with) the statistic of interest, Y, and assume that $\underline{C}$ has a *known* mean vector $\mathrm{E}[\underline{C}]$. The component variables of the vector $\underline{C}$ are the control variables.

Users of variance reduction techniques hope to more precisely estimate $\mathrm{E}[Y]$ by deriving a controlled statistic Y' that has less variance than Y. The idea behind control variables is to use the correlation between Y and $\underline{C}$ to exploit the knowledge about the expected

values of $\underline{C}$ to reduce the variance of Y'. A standard method for doing this is via the linear, additive combination of Y and the components of $\underline{C}$,

$$Y' = Y - \underline{\theta}^T (\underline{C} - \mathrm{E}[\underline{C}]). \tag{1}$$

The parameter vector $\underline{\theta}$ is a vector of q unconstrained constants that are to be chosen so as to minimize the variance of Y'. Note that some components of $\underline{C}$ may be ***known*** power transformations of other components, so that polynomial control schemes are included in formulation (1). Explicit expressions for the components of $\underline{\theta}$ that minimize the variance of Y' can be found in terms of the second-order moments of Y and $\underline{C}$, and with these parameters, Y' is an unbiased estimate of $\mathrm{E}[Y]$ in that $\mathrm{E}[Y'] = \mathrm{E}[Y]$.

In particular, consider the case of a single, additive, linear control

$$Y' = Y - \theta (C - \mathrm{E}[C]). \tag{2}$$

Here θ is chosen to minimize $\mathrm{Var}[Y']$. Assuming $\mathrm{Var}[Y] = \mathrm{Var}[C]$, it follows that

$$\begin{aligned} \mathrm{Var}[Y'] &= \mathrm{Var}[Y] + \theta^2 \mathrm{Var}[C] - 2\theta \mathrm{Cov}[Y, C] \\ &= \mathrm{Var}[Y] \left(1 + \theta^2 - 2\theta \rho[Y, C]\right). \end{aligned}$$

Differentiating with respect to θ and setting the resulting expression equal to zero yields the optimal value for θ i.e., the value that minimizes $\mathrm{Var}[Y']$:

$$\theta = \rho[Y, C],$$

where $\rho[Y, C]$ is the correlation between Y and C namely

$$\rho[Y, C] = \mathrm{Cov}[Y, C] / \sigma_Y \sigma_C. \tag{3}$$

Without the assumption of equal variances, it follows that the value of θ that maximizes the variance reduction is

$$\theta = \frac{\sigma_Y}{\sigma_C}\rho[Y,C], \tag{4}$$

where σ_Y represents the standard deviation of the random variable Y.

In the multiple control case when q is greater than one, it can be shown (see Kendall and Stuart, 1977, Chap. 27) that the values for $\underline{\theta}$ that minimize the variance of Y' are the multiple regression coefficients

$$\underline{\theta} = \left(\Sigma_{\underline{C}}^{-1}\right)\sigma_{Y,\underline{C}} \tag{5}$$

where $\Sigma_{\underline{C}}$ is the covariance matrix of $\underline{C}$ and $\sigma_{Y,\underline{C}}$ is the q-dimensional vector with components $\mathrm{Cov}[Y,C_j]$, for $j = 1,\ldots,q$. Rubinstein and Marcus (1985) demonstrated that the solution for $\underline{\theta}$ in the linear control of a single response, Y, is a special case of determining the canonical correlation coefficients for maximizing the correlation between linear combinations of multiple responses and multiple controls.

1. A Measure of the Effectiveness of a Control for Variance Reduction

One measure of effectiveness for a particular linear control is the percent variance reduction, a measure that involves the ratio of the variance of the controlled estimate Y' to the variance of the uncontrolled estimate Y. A high percent variance reduction implies that the control is effective at reducing the variance of the point estimate. For a single control, assuming the optimal value for θ in (4) is known, the percent variance reduction is

$$1 - \frac{\sigma_{Y'}^2}{\sigma_Y^2} = \rho^2(Y,C). \tag{6}$$

Equation (6) implies that for the control to be effective, one should choose a random variable which is "strongly" correlated with Y to be the control variable C.

For multiple controls, the percent variance reduction is the direct generalization

$$1 - \frac{\sigma_{Y'}^2}{\sigma_Y^2} = \mathcal{R}_{Y,\underline{C}}^2. \tag{7}$$

where

$$\mathcal{R}_{Y,\underline{C}}^2 = \frac{\sigma_{Y,\underline{C}}^T\left(\Sigma_{\underline{C}}^{-1}\right)\sigma_{Y,\underline{C}}}{\sigma_Y^2}$$

is the square of the multiple correlation coefficient between Y and $\underline{C}$. As before, the effectiveness of the control depends upon a large value for $R^2_{Y,\underline{C}}$. When the components of $\underline{C}$ are independent, one can simplify (7). For example, with two independent linear controls

$$1 - \frac{\sigma^2_{Y'}}{\sigma^2_Y} = \rho[Y, C_1]^2 + \rho[Y, C_2]^2 \,. \tag{8}$$

When the number of multiple controls to use is given, one should simply choose those controls which maximize the $\mathcal{R}^2_{Y,\underline{C}}$. However, determining the number of multiple controls to use is a more difficult problem which is complicated by the necessity of estimating the coefficients in $\underline{\theta}$.

2. Estimating the Coefficients for Linear Control of a Sample Mean

In the usual case in simulation, the values for θ or $\underline{\theta}$ must be estimated since not enough information is known about the joint distribution of Y and $\underline{C}$ to determine the regression coefficients in (5). For notation's sake, assume that one is using a single control. Assume that Y and C are sample means $\overline{Y}$ and $\overline{C}$ calculated in the usual manner from m i.i.d. replicates Y_i and C_i of random variables $\mathcal{Y}$ and $\mathcal{C}$ such that $\overline{Y} = (1/m)\sum_{i=1}^{m} Y_i$ and similarly for $\bar{C}$. One could generate sample estimates of the variance and covariances in (4) to estimate θ; however since θ is the coefficient of regression, an equivalent but computationally more convenient method for estimating θ is to use linear least-squares regression.

The regression coefficient θ can be estimated by the least-squares regression of $(C_i - \overline{Y})$ on $\theta(C_i - \mathrm{E}[C])$ using the regression model

$$(Y_i - \overline{Y}) = \theta(C_i - \mathrm{E}[C]) + \epsilon_i, \quad \text{for } i = 1, \ldots, m$$

where the C_i are considered fixed and ϵ_i is a mean-zero random variable independent of C_i.

3. The Loss Factor

In general, estimating the coefficients can cause a reduction in the percent variance reduction predicted by (6) or (7). Lavenberg, Moeller and Welch (1982) investigated the decrease in predicted variance reduction caused by using the individual samples to estimate $\underline{\theta}$ for a linear control of the sample mean. Under the assumption of multivariate normality

between the statistic of interest and the control, they concluded that the decrease in variance reduction due to estimating $\underline{\theta}$ could be predicted as a function of the number of independent samples of the statistic being controlled, m, and the number of controls whose coefficients had to be estimated, q. They used a loss factor of $(m-2)/(m-q-2)$ to predict the actual variance reduction as $\frac{\sigma^2_{Y'}}{\sigma^2_Y} = (1 - \mathcal{R}^2_{Y,\underline{C}})(m-2)/(m-q-2)$. The loss factor is a deterrent to adding more controls simply to achieve a small increase in the $\mathcal{R}^2(\cdot)$ in (7). As one selects more controls for a multiple control scheme, the impact of the loss factor can quickly overcome the benefits of increasing the $\mathcal{R}^2(\cdot)$. Thus one can not guarantee an improvement in the effectiveness of a linear control by simply adding more controls.

4. Measuring the Effectiveness of a Control at Reducing Sample Sizes

Lewis and Orav (1989, p. 262) mention an alternative measure for quantifying the effectiveness of a control scheme. They look at the square root of the ratio of the variance of the uncontrolled estimate to the variance of the controlled estimate i.e., $\sigma_Y/\sigma_{Y'}$. This ratio can be considered to be the ratio of the sample size that would be needed to achieve a given standard deviation without using the control scheme, to the sample size needed to achieve the same standard deviation using the control. When expressed in terms of the correlation coefficient for the controlled statistic and the control from (3), the ratio becomes $1/[1-\rho^2(\cdot)]^{1/2}$. Given a value for $\rho(\cdot)$, the formula gives the increase in the sample size that would be needed to achieve the same standard deviation without the control. Given a desired reduction in sample size, say 1/2, the formula implies that to achieve a given standard deviation while cutting the sample size in half, one must have $1-\rho^2(\cdot) = .25$, which implies a correlation coefficient of ± 0.86.

Linear controls are typically unable to reduce the sample size needed to attain a given σ_Y by as much as a half because the correlation between the statistic of interest and a linear function of the control variables is not high enough. Since many statistics have a nonlinear relationship with the control variables, one possible means for increasing the variance reduction for a given set of controls is to allow nonlinear transformations of the controls.

C. NONLINEAR CONTROLS

1. Definition of a Nonlinear Control

One can generalize the linear control scheme for q controls, (1), to include nonlinear transformations of random variables as controls for variance reduction. Let $h_j(C_j, \underline{\theta}_j)$, for $j = 1, \ldots, q$, be a transformation function of the random variable C_j and let $\underline{\theta}_j$ be a vector of coefficients where, depending upon $h_j(\cdot)$, the vector $\underline{\theta}_j$ may have more than one component. When incorporating nonlinear transformations of multiple controls, the linear control scheme (1) becomes

$$Y' = Y - \tilde{C} \tag{9}$$

$$= Y - H(\underline{C}, \underline{\theta}) \tag{10}$$

where, for our purposes, $H(\cdot,\cdot)$ is a linear additive combination of the q transformed controls, $h_j(C_j, \underline{\theta}_j)$, and their expected values, $\mathrm{E}\left[h_j(C_j, \underline{\theta}_j)\right]$, for $j = 1, \ldots, q$. The vector $\underline{\theta}$ contains the coefficients from the linear combination in addition to the q sets of coefficients from the individual transformations. $H(\underline{C}_j, \underline{\theta})$ will be referred to as the control function for $\tilde{C}$. If the control function has terms that are nonlinear in the unknown coefficients, $\tilde{C}$ will be said to be a nonlinear control. For ease of notation, the coefficients $\underline{\theta}$ may be suppressed in the expressions for $H(\cdot)$ and $h(\cdot)$. When there is only one control so that $q = 1$, the subscript j will be suppressed so that $h_j(\cdot) = h(\cdot)$.

In some simulations possible control variables may have low correlation with Y. For a given control, two of the possible sources for the low correlation between Y and C are:

1. there is in fact little structural relationship between Y and the control i.e., a bivariate scatter plot of Y versus C would look patternless, or
2. the structural relationship between Y and C is of a nonlinear form which is poorly approximated by a straight line (the linear regression line).

In the first case, a nonlinear control may or may not offer improvement over the linear control. In the second case, a nonlinear control can offer substantial improvement in variance reduction.

A simple example will show the potential benefits of nonlinear transformations. Let C be a Normal (0,1) random variable which is being used to control the sample mean of $Y = C^2$. It follows that

$$\mathrm{Cov}[Y, C] = \mathrm{E}\left[C^3\right] - \mathrm{E}\left[C^2\right] \mathrm{E}[C] = 0$$

so that $\rho(Y, C)$ is zero, which implies zero effectiveness for the linear control as well. Now allow the nonlinear transformation

$$h^*(C) = h(C, \theta) = C^\theta$$

with $\theta = 2$. The transformed random variable $h^*(C)$ is a χ_1^2 random variable with mean 1 and variance 2. It follows that

$$\mathrm{Cov}[Y, h^*(C)] = \mathrm{Var}\left[C^2\right] = 2 \Longrightarrow \rho[w, h^*(C)] = \frac{2}{2} = 1$$

so that the nonlinear control is completely effective. Therefore when evaluating a potential control, one should ask: *Can this random variable be transformed to have a "high" correlation with the statistic of interest while still having an analytically computable mean?*

2. Optimal Nonlinear Transformations and ACE

For some random variables, transformations do exist which will improve their correlation with Y.

- Let Y and $\underline{C}$, with q components C_j, for $j = 1, \ldots, q$, be random variables with a general but nonsingular joint distribution.
- Let $g(Y)$ and $h_j(C_j))$ for $j = 1, \ldots, q$, be mean-zero transformation functions of random variables Y and C_j such that $\mathrm{Var}[g(Y)] = 1$ and $\mathrm{Var}[h_j(C_j)] < \infty$, for $j = 1, \ldots, q$.

Breiman and Friedman (1985) proved the existence of optimal transformations for maximizing the correlation between $g(Y)$ and $H(\underline{C})$, a linear additive function of the mean-zero $h_j(C_j)$. The optimal transformation for a particular variable can be expressed in terms of the conditional expected values of given transformations of the other variables. In this bivariate case, where $H(\cdot) = h(\cdot)$ since $q = 1$, the pair of optimal transformations $g^*(\cdot)$ and

$h^*(\cdot)$ are:

$$g^*(Y) = \frac{\mathrm{E}[h^*(C) \mid Y]}{\|\mathrm{E}[h^*(C) \mid Y]\|}$$

and

$$h^*(C) = \mathrm{E}[g^*(Y) \mid C]$$

where $\|\cdot\| = \{\mathrm{E}[(\cdot)^2]\}^{1/2}$.

In the multiple control case, where $q > 1$,

$$g^*(Y) = \frac{\mathrm{E}\left[\sum_{j=1}^{q} h_j^*(C_j) \mid Y\right]}{\left\|\mathrm{E}\left[\sum_{j=1}^{q} h_j^*(C_j) \mid Y\right]\right\|} \tag{11}$$

and

$$h_j^*(C_j) = \mathrm{E}\left[g(Y) - \sum_{k \neq j} h_k^*(C_k)\right]. \tag{12}$$

The transformations $g^*(\cdot)$ and $h^*(\cdot)$ in (11) and (12) will usually be nonlinear, the exception being when Y and $\underline{C}$ have a multivariate normal distribution.

Results from Lancaster (1966) can be used to show that if Y and $\underline{C}$ have a multivariate normal distribution, the solutions for $g(Y)$ and $H(\underline{C})$ which have maximal correlation between $g(Y)$ and $H(\underline{C})$, over all measurable functions of finite variance, are the linear transformations which yield the first Hotelling canonical variables. In other words, when Y and $\underline{C}$ have a multivariate normal distribution, using the linear control scheme (1), with the multiple regression coefficients for $\underline{\theta}$ from (5), produces the greatest amount of variance reduction. Conversely, whenever the joint distribution of Y and $\underline{C}$ is not multivariate normal, a nonlinear control offers the possibility for greater variance reduction over a linear control.

D. APPROXIMATING OPTIMAL NONLINEAR TRANSFORMATIONS FOR NONLINEAR CONTROLS

1. Estimating the Optimal Nonlinear Transformations

Determining the optimal transformations in (11) and (12) analytically requires the joint distribution of Y and $\underline{C}$ which, in the context of a simulation, is unknown. In the multivariate normal case, the form of the transformations are known to be linear and one can estimate the coefficients using linear least-squares regression. With a nonlinear control, one must first estimate the form of the transformations.

Breiman and Friedman (1985) also developed the Alternating Conditional Expectation Algorithm (ACE) as a means for generating nonparametric estimates of the optimal transformations (11) and (12). In the ACE implementation for finite data sets of continuous variables, data smooths are used in place of the analytical conditional expected values. The ACE algorithm produces estimates of the optimal transformations as sets of fitted values, one set for each variable. Plotting the fitted values against the original values gives the shape of the estimated transformation for each variable.

ACE also provides an estimate of the maximum obtainable squared correlation between the transformed response and the sum of transformed predictors. Given a data set of n samples of Y and $\underline{C}$, and a set of transformations $g(Y)$ and $h_j(C_j)$ for $j = 1, \ldots, q$, the R^2 estimate is calculated as 1 minus the sample mean-squared error, or

$$R^2 = 1 - \frac{1}{n} \sum_{i=1}^{n} \left[g(Y_i) - \sum_{j=1}^{q} h_j(C_{j,i}) \right]^2 .$$

This R^2 estimate is quite useful as it provides an estimate of an upper bound on the percent variance reduction that one can obtain using the given set of controls. Thus given a set of linear controls, one can use ACE to determine if the use of nonlinear transformations of the controls would improve the percent variance reduction. One technique is to use ACE on data from a small test simulation for various potential controls and then compare the estimates of R^2 to help select the controls.

Since ACE does not supply any parametric clue to the optimal transformations of the individual components of $\underline{C}$, approximations are needed for these transformations. A desirable feature for the approximations is that they contain the linear additive case (1)

as a special set of parameter values, thus ensuring that one attains at least the known variance reduction for this case. The approximations studied in the next section take two forms, piecewise linear controls, and standard statistical parametric transformations, used separately or conjointly on each component of $\tilde{C}$. It should be emphasized again that an additional constraint on an approximating transformation $g(C)$ is that for $g(C)$ to be usable as a nonlinear control, one must be able to calculate the mean of $g(C)$.

2. Piecewise Transformations of Controls

Statistics from simulations are often nonlinear functions of the input random variables from which they are derived. Therefore one might expect some nonlinear controls to have a higher correlation with Y than linear controls. Given an initial guess at a viable linear control, one type of nonlinear control can be formed by using indicator functions and "cutpoints" to form piecewise transformations of the control.

For example, a control variable X is split into two control variables C_1 and C_2 by using a fixed cutpoint δ and functions $h_j(X, \theta_j)$ as follows:

$$C_j = \mathrm{I}_j(X) h_i(X, \theta_j) \quad \text{for } j = 1, 2, \tag{13}$$

where

$$\mathrm{I}_1(X) = \begin{cases} 1 & \text{if } X \le \delta; \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \mathrm{I}_2(X) = \begin{cases} 1 & \text{if } X > \delta; \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

By judicious choice of a fixed value for the cutpoint δ or perhaps multiple cutpoints, least-squares multiple regression can achieve a better fit without the use of additional control variables. As an example let X be distributed as an independent Uniform (-.5, .5) variate. Let $Y = X^2 + \epsilon$ where ϵ is distributed as independent Normal (0, 0.01). In a small simulation of 300 samples, using just X as a linear control as in (1), linear least-squares regression yielded an R^2 of 0.00. However using X to form two new controls as in (13), where $h_j(X, \theta_j$ was linear both in X and θ_j for $j = 1, 2$ and $\delta = 0$, yielded an R^2 of .92 using linear least-squares regression.

If the functions $h_i(\cdot)$ are linear, then the linear control is a special case of this transformation. If one lets $\delta \to \infty$ or $\delta \to -\infty$, the ordinary linear control is obtained.

Of course, care must be taken in determining the form of the control function to ensure it has mean zero i.e., $E[C_1]$ and $E[C_2]$ must be known. Note also that the regression is still linear if δ is *given*, but it is nonlinear otherwise. Finding an optimal δ then becomes, in general, a discontinuous nonlinear, mathematical programming problem. An alternative method for finding δ is to use graphical analysis of the relationship between Y and C in selecting the initial cutpoint(s).

3. Transformations of Controls

Several standard transformations are used in statistics and data analysis (Chambers et.al., 1983) and these can be applied as approximations for the optimal transformation of a control variable C. Power transformations of controls, in addition to piecewise transformations of controls, introduce nonlinearity into the controlled estimate of $E[Y]$ while containing the untransformed control as a special case. The power transformation used initially in this study is of the well-known form

$$Z = (X^p - 1)/p, \quad \text{for } p > -1. \tag{15}$$

This scaled power transformation has the property that as $p \to 0$ the limit is $\ln X$ and when $p = 1$ it gives a shifted version of the original variable.

This power transformation (15) can have vastly different effects for $X > 1$ and $X < 1$. The curves in Figure 1 represent a sample of possible transformations. As one increases p, the change in the nature of the function on either side of $X = 1$ becomes more drastic. For large values of p, large values of X are given added weight while for small values of p, the small values of X are given the additional weight. Note that when $p = 1$, this is simply the linear transformation. Thus optimizing using this transformation assures variance reduction at least as good as in the linear case.

Using, for example, the single control variable C, the resulting nonlinear control function is

$$\tilde{C} = \theta \left\{ \frac{C^p - 1}{p} - E\left[\frac{C^p - 1}{p}\right] \right\},$$

which has two parameters, p and θ. Of course, combinations of piecewise transformations and power transformations are also possible by letting the $h_i(\cdot)$ in (13) be nonlinear

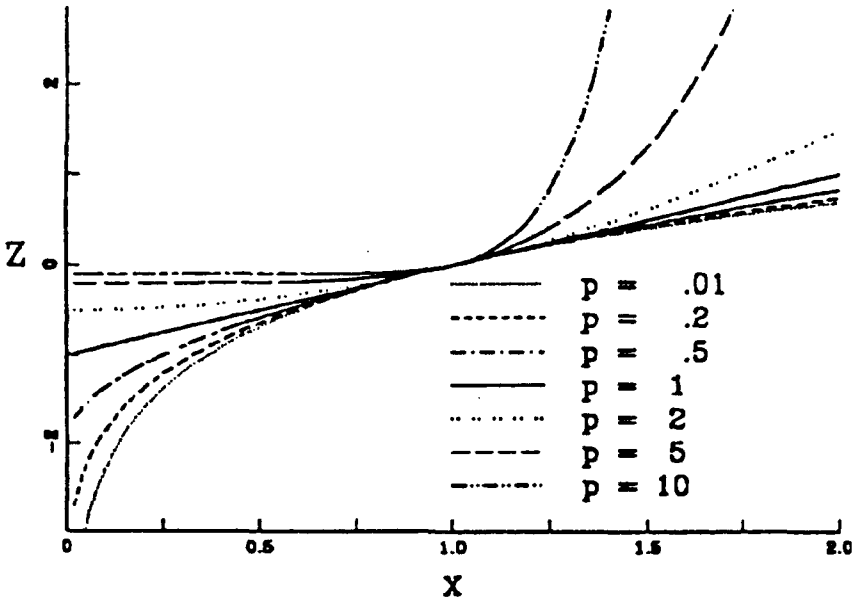


Figure 1. Examples of power transformations of a variable X.

functions, and it is this combination of nonlinear controls that is the main thrust of the example to follow. With this combination one hopes to come close to the maximum theoretical variance reduction which could be obtained.

Other transformations one might use for inducing nonlinearity include

1. $Z = \left(e^{\gamma X} - 1\right)/\gamma$,
2. $Z = \left((X^p - 1)\left(e^{\gamma X} - 1\right)\right)/p\gamma$, or
3. $Z = \left(e^{\gamma(X^p-1)/p} - 1\right)/\gamma$.

These transformations represent a broad spectrum of transformations on a variable as can be seen in Figures 2, 3 and 4. Note also that transformation 1 and transformation 3 contain the linear case as a special set of parameter values. Transformation 1, is a positive weighting of all values, with large values weighted more than small values. By varying the γ parameter, one can scale the effects of the weights from very large for large γ to very slight for small γ. Transformation 2, applies small negative weights for values less than 1. For values larger than 1 it allows for a wide range of positive weighting schemes as in transformation 1. third Transformation 3, is similar to the straightforward power transformation, (Figure 1),

but with more parameters. Thus it allows for more flexibility and increased curvature for smaller values of the parameters. The difficult part with these transformations, as usual, is computing the necessary expected values.

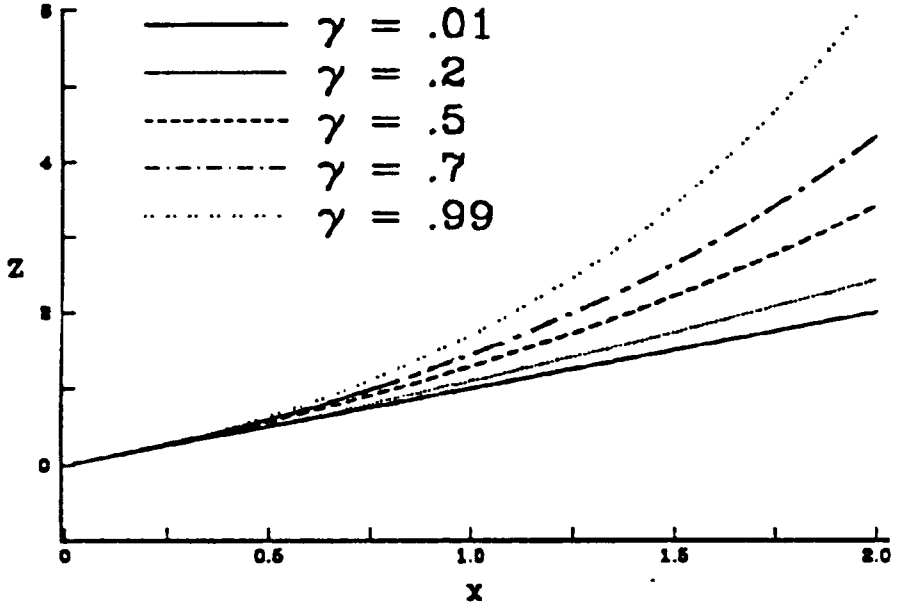


Figure 2. Transformation 1 applied to a variable X.

E. THE ROLE OF NONLINEAR LEAST-SQUARES REGRESSION

Given that one has selected appropriate nonlinear transformations for the components of $\underline{C}$, the optimal values for the parameters of the transformations can be obtained by minimizing the variance of Y'. Unfortunately, the results are not explicit functions of the joint and higher moments between Y and $\underline{C}$ as they are for the linear controls. Starting with (9), one can write

$$\begin{aligned} \frac{\mathrm{Var}[Y']}{\mathrm{Var}[Y]} &= 1 + \frac{\mathrm{Var}\left[\tilde{C}\right]}{\mathrm{Var}[Y]} - 2\frac{\sigma_{\tilde{C}}}{\sigma_Y}\rho\left[Y, \tilde{C}\right] \\ &= 1 + k^2 - 2k\rho\left[Y, \tilde{C}\right] \end{aligned} \tag{16}$$

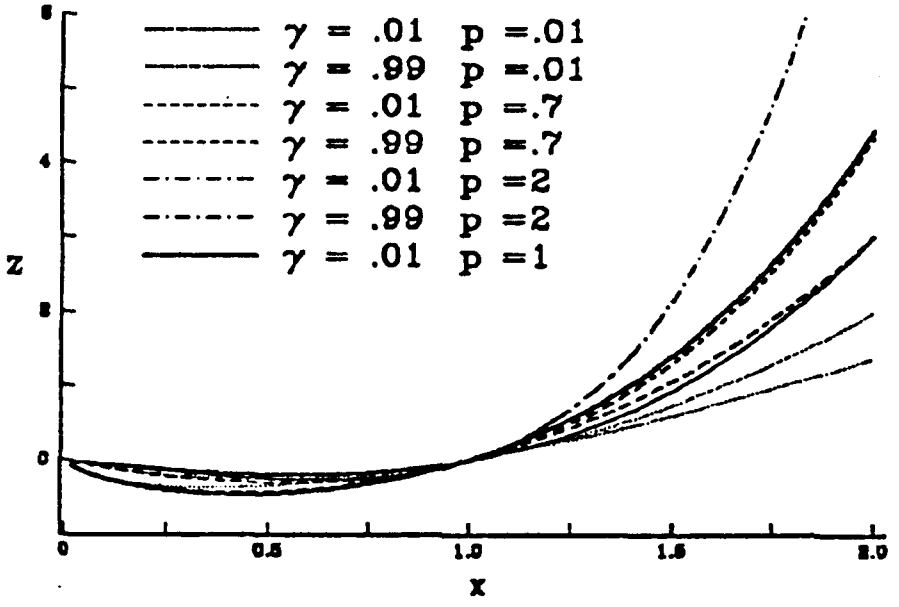


Figure 3. Transformation 2 applied to a variable X.

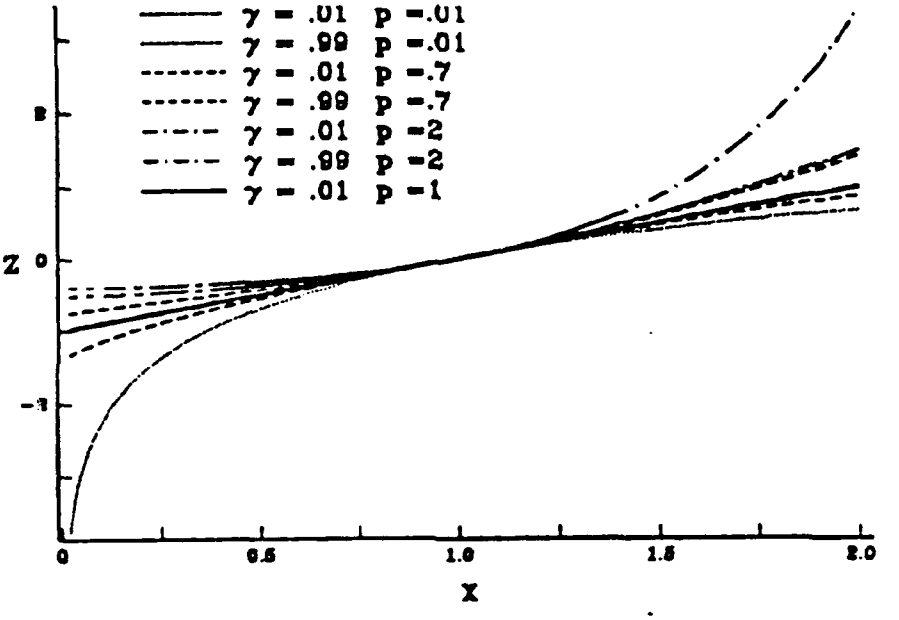


Figure 4. Transformation 3 applied to a variable X.

and

$$1 - \frac{\mathrm{Var}[Y']}{\mathrm{Var}[Y]} = 2k\rho\left[Y, \tilde{C}\right] - k^2, \tag{17}$$

where k is positive valued. While this last equation is simple in form, both $\rho\left[Y, \tilde{C}\right]$ and $k = \sigma_{\tilde{C}}/\sigma_Y$ are functions of the unknown parameters in $\tilde{C}$.

In practice, one has insufficient theoretical information about the joint probability properties of Y and the components of $\underline{C}$ to determine the parameters in $\tilde{C}$. Instead one has a simulation sample of size m of independent replications, $\{Y_i, \underline{C}_i\}$ for $i = 1, \ldots, m$, from which to estimate E[Y]. Regardless of whether the sample is large or small i.e., is a pilot sample or all of the simulation data that will be available, one wants to minimize the sample variance of Y'. Minimizing the sample variance involves, after subtracting $\overline{Y}$ from both sides of (9), minimizing

$$\frac{\sum_{i=1}^{m}\left(Y_i' - \overline{Y}\right)^2}{m} = \frac{\sum_{i=1}^{m}\left(Y_i - \overline{Y} - \tilde{C}_i\right)^2}{m} \tag{18}$$

$$= \frac{\sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)^2}{m} + \frac{\sum_{i=1}^{m}\tilde{C}_i^2}{m} - \frac{2\sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)\tilde{C}_i}{m} \tag{19}$$

The left-hand side of (18) is the quantity to be minimized as $\mathrm{E}\left[\overline{Y}\right] = \mathrm{E}\left[\overline{Y'}\right] = \mathrm{E}[Y]$ since $\mathrm{E}\left[\tilde{C}\right]$ is known to be zero. Thus either $\overline{Y}$ or $\overline{Y'}$ can be used in the estimate of the variance of Y'. Equation (18) shows that this estimate of the variance of Y' is equal to the residual sum of squares of the least-squares regression of $Y - \overline{Y}$ on $\tilde{C}$. Equation (19) involves, in its first term, the total sum of squares, which estimates the variance of Y; in its second term the sample variance of the zero-mean $\tilde{C}$; and in the last term the sample covariance of Y and $\tilde{C}$. Rearranging terms in (19), one gets

$$\frac{\sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)^2}{m} - \frac{\sum_{i=1}^{m}\left(Y_i' - \overline{Y}\right)^2}{m} = \frac{2\sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)\tilde{C}_i}{m} - \frac{\sum_{i=1}^{m}\tilde{C}_i^2}{m}$$

or

$$\frac{\sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)^2 - \sum_{i=1}^{m}\left(Y_i' - \overline{Y}\right)^2}{\sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)^2} = \frac{2\sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)\tilde{C}_i}{\sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)^2} - \frac{\sum_{i=1}^{m}\tilde{C}_i^2}{\sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)^2}. \tag{20}$$

The left-hand side of (20) is the usual R^2 regression measure and equation (20) may be rewritten as

$$R^2 = 2\frac{s_{\tilde{C}}}{s_Y} r\left[Y,\tilde{C}\right] - \frac{s^2_{\tilde{C}}}{s^2_Y} = 2\hat{k}\, r\left[Y,\tilde{C}\right] - \hat{k}^2 \qquad (21)$$

where $s^2_{\tilde{C}}$ represents the sample variance and $r\left[Y,\tilde{C}\right]$ represents the sample correlation coefficient.

As the sample analog to (17), (21) indicates that maximizing R^2 through nonlinear least-squares regression is equivalent to maximizing sample variance reduction when the optimal parameters are unknown. Thus for $\underline{C}$ with multiple components, maximizing the effectiveness of $\tilde{C}$ can be accomplished through estimating the parameters of $\tilde{C}$ via multiple least-squares regression of $Y - \overline{Y}$ on $\tilde{C}$. A similar result relating optimal regression and optimal correlation can be found in the ACE article (Breiman and Friedman, 1985).

With linear controls, linear least-squares regression will provide a global minimum for the residual sum of squares, in turn maximizing the variance reduction for the sample. When the control function is nonlinear, nonlinear least-squares regression will not necessarily determine parameter values that globally minimize the residual sum of squares. With a nonlinear control function $\tilde{C}$, there may be many suboptimal local minima. In this case the choice of initial values for the parameters $\underline{\theta}$ in the nonlinear regression may significantly affect the amount of variance reduction obtained. If one uses as starting values for $\underline{\theta}$ the special values that represent the linear case for the control, one should always do at least as well as the linear case.

One must be careful that while multiple regression may be computationally useful, the distribution theory behind multiple regression, which assumes fixed independent variables, does not apply. Consequently, although the estimates of $\underline{\theta}$ are identical, (see Sampson, 1974) confidence intervals on parameter estimates cannot be determined directly from the regression results.

F. AN EXAMPLE OF USING NONLINEAR CONTROLS

Estimating the mean of the Anderson-Darling goodness-of-fit statistic, W^2_n, (Anderson and Darling, 1952) provides a good example of the benefits of piecewise internal controls

and power transformations. The example is artificial since $\mathrm{E}[W_n^2]$ is known to be one for all n. However, it is useful as an illustration.

The statistic W_n^2 can be determined as a function of n independent unit exponential random variables E_j for $j = 1, \ldots, n$ (Lewis and Orav, 1989, p. 369). Note first that one can write W_n^2 as a function of order statistics from a unit exponential distribution as follows:

$$W_n^2 = -n - \left(n^{-1}\right) \sum_{i=1}^{n} \left[-(2i-1) \ln \left\{ 1 - e^{-\tilde{E}_{(i)}} \right\} + \{2(n-i)+1\} \tilde{E}_{(i)} \right] \qquad (22)$$

where the $\tilde{E}_{(i)}$ are the order statistics from a unit exponential population. One method for generating a sample of the order statistics from a unit exponential problem is to simply order n i.i.d. realizations from a unit exponential population. A second method, which is used here, generates the sample of order statistics from a unit exponential population, $\tilde{E}_{(1)}, \ldots, \tilde{E}_{(n)}$, from n independent unit exponential random variables, E_j, for $j = 1, \ldots, n$, as follows:

$$\tilde{E}_{(i)} = \sum_{j=1}^{i} \frac{E_j}{(n-j+1)} \quad \text{for } i = 1, \ldots, n. \qquad (23)$$

Note that the order statistics $\tilde{E}_{(1)}, \ldots, \tilde{E}_{(n)}$ produced by (23) are *not* the order statistics of the original sample of n independent exponentials $\{E_1, \ldots, E_n\}$. However, they are a sample of the order statistics from an n-sized sample of i.i.d. unit exponential random variables. Together (22) and (23) give W_n^2 as a function of n independent exponential random variables. The independence of these random variables makes them ideal for controlling W_n^2. The case $n = 2$ is presented here, for which (8) holds with $C_1 = E_1$ and $C_2 = E_2$.

As mentioned before, graphical methods can sometimes be useful in determining types of controls or aspects of controls. Two useful plots of W_2^2 are presented here. Figure 5 is a surface plot of W_2^2 over a small region of the E_1, E_2 plane where the majority of values occur. This is not a density plot, but a representation of the functional relationship between the two independent exponentials and the W_2^2 value each pair generates. Subsequent surface plots of the control functions likewise do not portray density; just the surface generated by the control function. As an indicator of the density of points on the W_2^2 surface, Figure 6 is a sample bivariate histogram of 1000 independent pairs of unit exponentials. While one could

plot an actual bivariate exponential density, the discrete nature of the histogram allows easier visual comparisons of density. Together, Figures 5 and 6 indicate why nonlinear controls may prove useful for controlling W_2^2. Clearly the relationship between W_2^2, E_1 and E_2 is highly nonlinear, suggesting the use of nonlinear controls. Figure 6 supports one's intuition that the majority of pairs of the bivariate exponential are close to the origin. Suspecting this, one may be tempted to use a linear control to just approximate the surface in this region. However, Figure 6 also shows a significant number of pairs throughout the plane. Thus in order for $\tilde{C}$ to be an effective control, the entire surface should be approximated by the control. This would require a nonlinear control and nonlinear regression.

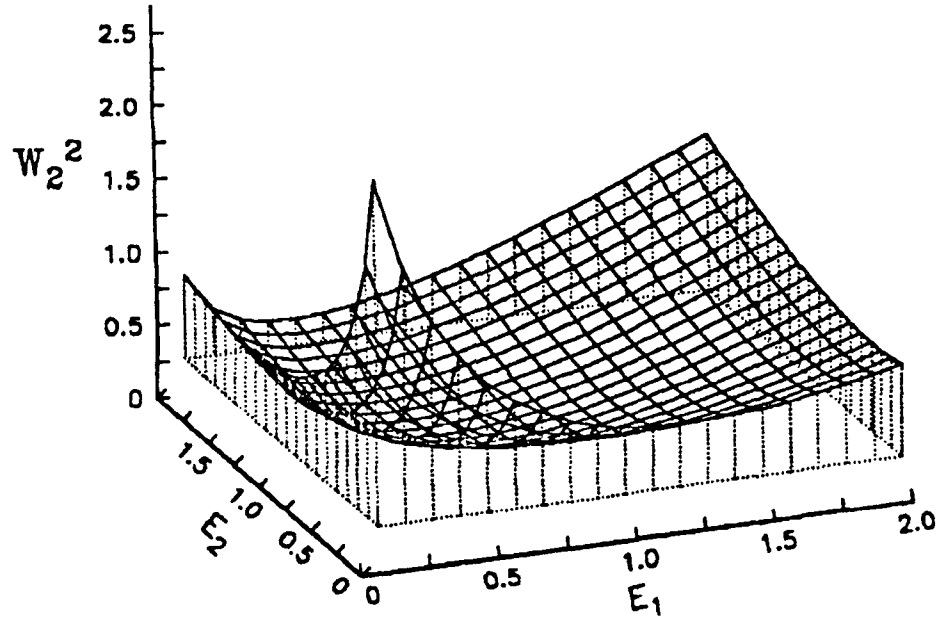


Figure 5. The nonlinear surface of W_2^2 as a function of two variables E_1 and E_2.

Six different linear and nonlinear control functions for estimating the mean of W_2^2 were evaluated using a single sample of 500 pairs of unit exponentials and their associated W_2^2 values. The first five control functions are:

$$\tilde{C} = \theta_1 (E_1 - 1) + \theta_2 (E_2 - 1); \tag{24}$$

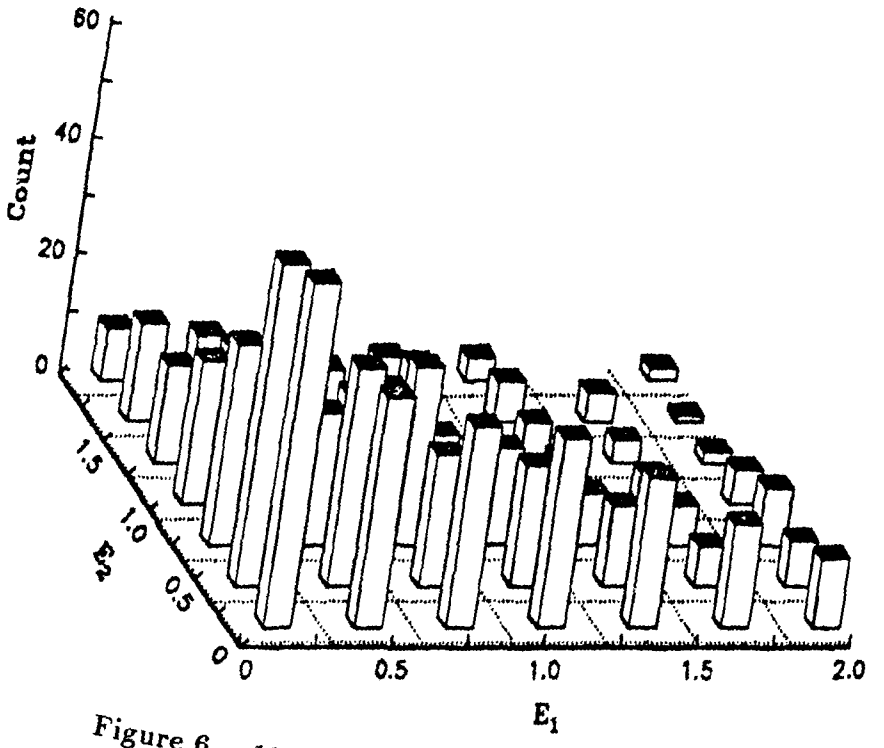


Figure 6. 1000 Samples of E_1, E_2 pairs

$$\tilde{C} = \theta_1 (E_1 - 1) + \theta_2 (E_2 - 1) + \theta_3 \left(E_1^2 - 2\right) + \theta_4 \left(E_2^2 - 2\right);$$

$$\tilde{C} = \sum_{j=1}^{2} \theta_j \left[\frac{E_j^{p_j} - 1}{p_j} - \mathrm{E}\left[\frac{E_j^{p_j} - 1}{p_j} \right] \right]; \tag{25}$$

$$\begin{aligned} \tilde{C} \;=\; & \theta_1 (E_1 - 1) + \theta_2 (E_2 - 1) \\ & + \theta_3 \left(\frac{E_1^{p_1} - 1}{p_1} - \mathrm{E}\left[\frac{E_1^{p_1} - 1}{p_1} \right] \right) + \theta_4 \left(\frac{E_2^{p_2} - 1}{p_2} - \mathrm{E}\left[\frac{E_2^{p_2} - 1}{p_2} \right] \right); \end{aligned} \tag{26}$$

and

$$\tilde{C} = \sum_{j=1}^{2} \sum_{k=1}^{2} \mathrm{I}_k (E_j) \theta_{jk} \left\{ \frac{E_j^{p_{jk}} - 1}{p_{jk}} - \mathrm{E}\left[\frac{E_j^{p_{jk}} - 1}{p_{jk}} \right] \right\}, \tag{27}$$

where in (28)

$$\mathrm{I}_1 (E_j) = \begin{cases} 1 & E_j \le \delta \\ 0 & \text{otherwise,} \end{cases} \quad \text{and } \mathrm{I}_2 (E_j) = \begin{cases} 1 & \delta < E_j \\ 0 & \text{otherwise} \end{cases} \quad j = 1, 2. \tag{28}$$

The sixth and final control function uses two cutpoints and can be written as

$$\tilde{C} = \sum_{j=1}^{2} \sum_{k=1}^{3} \mathrm{I}_k(E_j)\theta_{jk} \left\{ \frac{E_j^{p_{jk}} - 1}{p_{jk}} - \mathrm{E}\left[\frac{E_j^{p_{jk}} - 1}{p_{jk}} \right] \right\}, \tag{29}$$

where

$$\mathrm{I}_1(E_j) = \begin{cases} 1 & E_j \le \delta_1 \\ 0 & \text{otherwise,} \end{cases} \qquad \mathrm{I}_2(E_j) = \begin{cases} 1 & \delta_1 < E_j \le \delta_2 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathrm{I}_3(E_j) = \begin{cases} 1 & E_j > \delta_2 \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } j = 1,2.$$

As expected, the simplest controls, (24) and (25), with straightforward control functions, were usually less effective than the nonlinear controls. The controls given by (24) and (25) are referred to as the "standard" controls because their unknown parameters can be computed using linear least-squares regression. Since the necessary expected values of the controls just involved the first two moments of the exponentials, they were determined analytically and not estimated. The remaining parameters for controls (24) and (25), respectively θ_1 and θ_2, and θ_1, θ_2, θ_3 and θ_4 were computed using linear least-squares regression.

Since control (24) is a linear function of E_1 and E_2 and W_2^2 is a very nonlinear function of E_1 and E_2, this control, not surprisingly, achieved an R^2 of only .2265. This poor performance could also be predicted by using the sample estimates for $\rho[W_2^2, E_1]$ and $\rho[W_2^2, E_2]$ in (8). If the estimates were the true correlations, the optimal θ's would only yield a 22.66 percent variance reduction. The parabolic shape of (25), as shown in Figure 7, enabled the control function to achieve an R^2 of .5627. While better, as will be seen shortly from the ACE results, it is far from optimal. Note that on the graphs demonstrating the controls, the predicted values of the controls are centered about zero, the mean of $\tilde{C}$.

For control (27) only the linear terms' expected values could be calculated analytically. The other two expected values were functions of the unknown parameters and had to be recalculated based on the current parameters during the optimization. For controls (26),

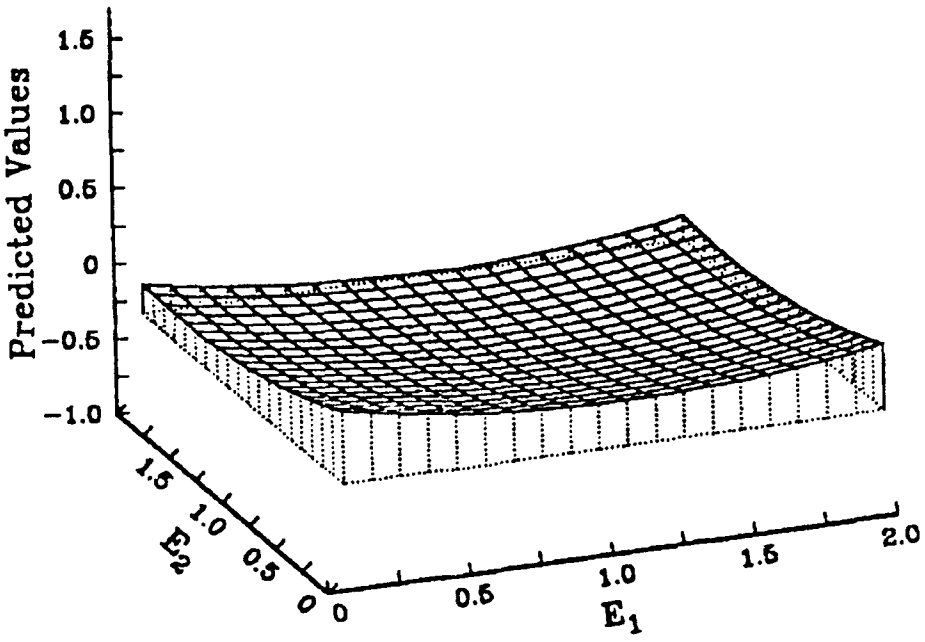


Figure 7. Surface generated by the parabolic linear control given in (24). The control is linear since the powers are fixed.

(28), and (29) none of the expected values could be determined analytically so all were calculated during the optimization. All of the parameters for the nonlinear controls were estimated via the nonlinear regression segment of IBM's experimental APL2-based GRAFSTAT statistical graphics package. For nonlinear regression, GRAFSTAT uses a form of the Marquadt algorithm (Marquadt, 1963) which allows bounds to be placed on the parameters (see Bard, 1974). Lower bounds of -.99 were necessary on the power parameters, p_{jk}, since the expected values of the exponentials (involving the gar function) are not defined for $p_{jk} \leq -1$. A reasonable upper bound on each p_{jk} was found useful in speeding convergence.

As the control functions became more nonlinear, their effectiveness usually increased. Allowing the powers to float in control (26) versus being fixed in control (24) gave a slight improvement; the R^2 went from .2265 up to .4640. This was not as good however as the "standard" control (25) with two linear terms and two quadratic terms which achieved an R^2 of .5627. Adding the two linear terms to control (26) resulted in control (27). Now allowing the powers to float in control (27), versus being fixed in control (25), enabled the

surface to fit more closely and thus the R^2 for (27) was .7422. This definite improvement over the "standard" controls can be seen in Figure 8.

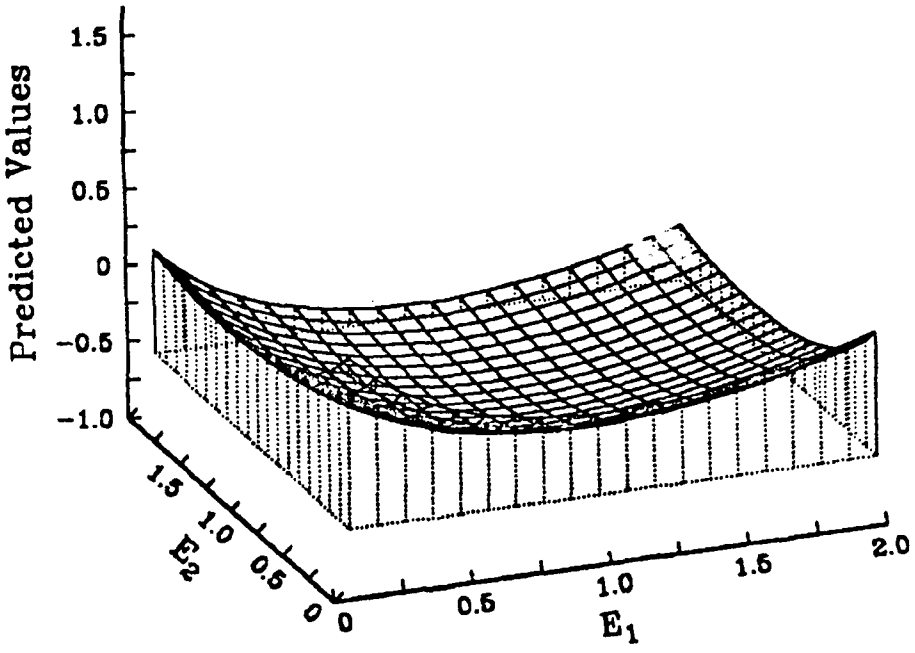


Figure 8. Surface generated by "non-standard" control with linear and nonlinear terms given in (26).

Originally, the cutpoints for controls (28) and (29) were parameters to be optimized. Unfortunately, this made the optimization unstable and the results unreliable. Thus, the cutpoints were fixed at selected quantiles and not included as parameters in the nonlinear regression. Selection of a good cutpoint was done by examining the results of a short sequence of regressions. For control (28) a cutpoint at the .5 quantile was the most effective one found for this sample. Comparing Figure 9 to Figure 8 shows the impact of adding nonlinearity by the use of the cutpoint. The R^2 for control (28) was .8216. The results of using the estimated parameters for (28) on three independent samples of 1000, Table 1, indicate that even though the regression-estimated parameters are biased for the original sample, (28) is still effective in controlling other samples.

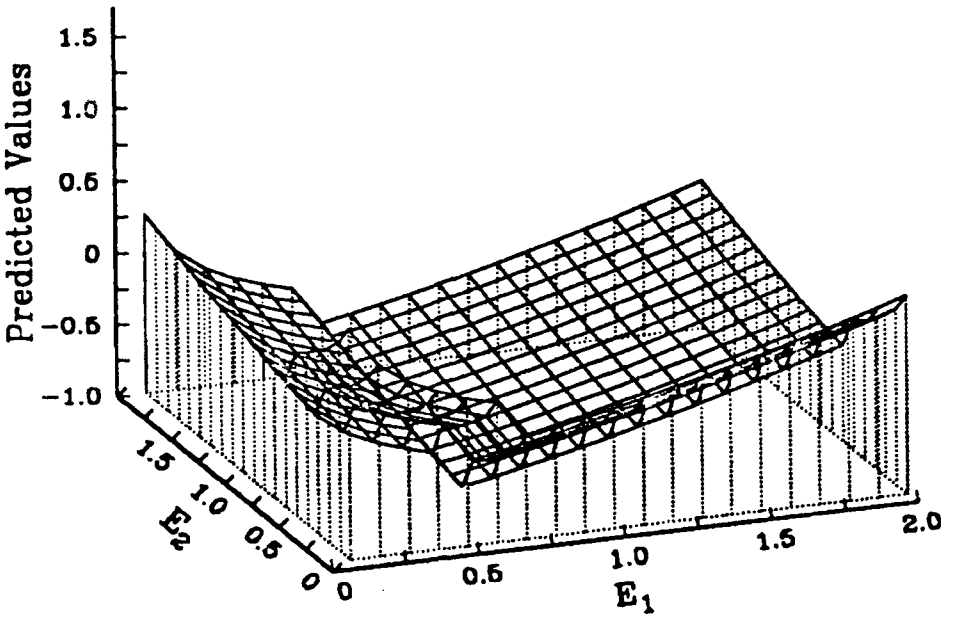


Figure 9. Surface generated by the nonlinear, single-cutpoint control given in (27).

	Sample 1	Sample 2	Sample 3
$\overline{W_2^2}$	.9882	1.0022	1.0219
$s_{\overline{W_2^2}}$	.0261	.0262	.0282
$\overline{W_2^{2\prime}}$	.9972	1.0095	1.0238
$s_{\overline{W_2^{2\prime}}}$	.0110	.0124	.0129
R^2	.8239	.7759	.7905

TABLE 1. Effect of the nonlinear, single-cutpoint control given in (27) on three independent samples other than the regression sample.

As the number of cutpoints increases to two for control (29), one gets a more effective control at the cost of increased computational complexity. The computational complexity increases because the additional cutpoint creates more parameters and because the computation of expected values becomes more expensive. As before, the cutpoints were fixed at selected quantile values. Which values to select was a matter of performing a series of regressions on a grid of values. Figure 10 shows that some pairs of cutpoints were better than others. Figure 11 shows that the best cutpoints for this sample on the grid examined, the .30 and .65 quantiles, yield a control that is an excellent approximation to the W_2^2 surface. The regression with these cutpoints on the original sample yielded an R^2 of .8372.

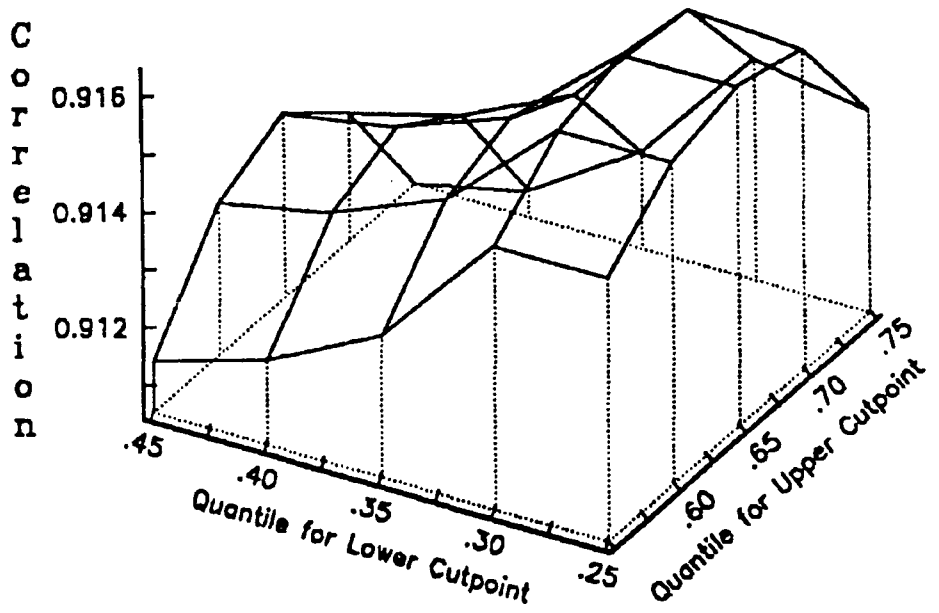


Figure 10. Effects of changing the cutpoints on correlation

This last control, (29), was then tested on independent samples and the R^2 was compared to results from ACE. Table 2 indicates the results for three samples of 1000 W_2^2 values. Again the R^2 values are almost as good as the original sample, and the control is effective in all three cases. ACE was given the data generated by using the cutpoints on the original sample as the independent variables. The R^2 value derived by ACE was .8560 showing that control (29) is nearly optimal for the control variables used.

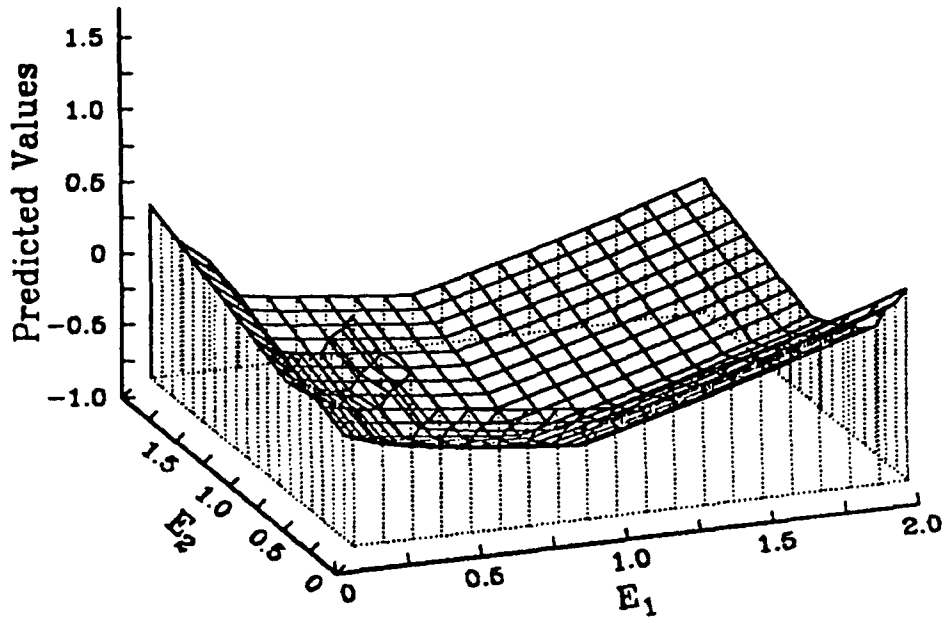


Figure 11. Surface generated by the nonlinear, double-cutpoint control given in (28).

	Sample 1	Sample 2	Sample 3
$\overline{W_2^2}$	.9534	1.0230	.9842
$s_{\overline{W_2^2}}$	.0230	.0265	.0255
$\overline{W_2^{2'}}$	.9997	1.0157	1.0001
$s_{\overline{W_2^{2'}}}$	.0094	.0121	.0110
R^2	.8350	.7925	.8153

TABLE 2. Effect of the nonlinear, double-cutpoint control given in (28) on three independent samples other than the regression sample.

G. SUMMARY

Linear controls are a well-known technique for variance reduction in computer simulation. Unfortunately, the relationship between the statistic of interest and the control is often poorly approximated by a straight line so the control has limited effectiveness. Nonlinear controls are a natural remedy for improving the effectiveness of a given set of controls. At the cost of greater analytical and computational effort in deriving the control, one can use the increased effectiveness of the control to reduce the sample size needed to achieve a given precision, thereby saving resources from the simulation.

In addition to controlling the estimate of a mean, one can also use nonlinear controls on the more difficult problem of quantile estimation in simulation. The next chapter discusses the use of nonlinear controls for quantile estimation where an important concern is also obtaining a reliable estimate of the variance of the controlled estimate.

III. VARIANCE REDUCTION FOR QUANTILE ESTIMATES IN SIMULATIONS VIA NONLINEAR CONTROLS

A. INTRODUCTION

As remarked in Chapter I, estimation and simulation of quantiles is different from the case of quantifiers that can be estimated as means. This chapter begins with a short discussion of quantiles and the properties of a quantile estimator, with emphasis on the need for a reliable estimator for the variance of the quantile estimator. The next section discusses linear controls for quantile estimates and the subtleties involved with estimating the coefficients for the control functions. The discussion of linear controls is followed by a discussion of the application of nonlinear controls to reducing the variance of quantile estimates for a fixed simulation sample size. The final part of the chapter presents an extract of results from a simulation experiment where crude, linearly controlled and nonlinearly controlled estimators are compared. Throughout, the emphasis is on quantile estimation for continuous random variables, though other cases are of interest.

B. QUANTILES

1. Properties of a Quantile Estimator

Let Y be a random variable with a right-continuous distribution function defined by

$$F_Y(y) = \Pr\{Y \leq y\}, \quad -\infty < y < \infty.$$

Following Serfling (1980) define the α quantile of Y, y_α, for $0 < \alpha < 1$, as the value

$$y_\alpha = F_Y^{-1}(\alpha) = \inf\{y : F_Y(y) \geq \alpha\}. \tag{30}$$

If $F_Y(y)$ is strictly increasing, y_α is unique for each α. Additional restrictions on $F_Y(y)$, such as continuity at y_α, may be needed for the existence of certain asymptotic properties and will be stated as required.

Given a simulation sample of n independent and identically distributed (i.i.d.) samples of Y, namely $Y_1, \ldots, Y_n$, one can construct a sample distribution function, F_n, by placing at each observation Y_i, a probability mass $1/n$. Thus F_n may be represented as

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(Y_i \leq y), \quad -\infty < y < \infty$$

where $\mathrm{I}(\cdot)$ is an indicator function which returns 1 if the argument is true and 0 otherwise.

For a sample of size n, one can define a nonparametric estimator of the α quantile, $\hat{y}_\alpha(n)$, as the sample α quantile of the sample distribution function, or

$$\hat{y}_\alpha(n) = F_n^{-1}(\alpha).$$

Using the sample α quantile to estimate y_α is equivalent to using the order statistics of the sample, $Y_{(1)} <, \ldots, < Y_{(n)}$, and defining a nonparametric estimator of the α quantile, $\hat{y}_\alpha(n)$, as in Lewis and Orav (1989), as

$$\hat{y}_\alpha(n) = Y_{(r)} = \begin{cases} Y_{(n\alpha)} & \text{if } n\alpha \text{ is an integer} \\ Y_{(\lfloor n\alpha \rfloor + 1)} & \text{if } n\alpha \text{ is not an integer} \end{cases} \tag{31}$$

where $\lfloor w \rfloor$ denotes the integral part of w.

For a given n and α, the quantile estimator $\hat{y}_\alpha(n)$ is the rth order statistic from the n-sized sample where r is determined as in (31). The following results on the distribution of $\hat{y}_\alpha(n)$ are well known (David, 1970, Chap. 1-3 or Kendall and Stuart, 1977, pp. 251-252).

Let $F_{\hat{y}_\alpha(n)}(y)$ be the cumulative distribution function of the quantile estimator. Then $F_{\hat{y}_\alpha(n)}(y)$ can be written as

$$\begin{aligned} F_{\hat{y}_\alpha(n)}(y) &= \Pr\{\hat{y}_\alpha(n) \leq y\} \\ &= \Pr\{\text{ at least } r \text{ of the } n \; Y_i \text{ are} \leq y\} \\ &= \sum_{i=r}^{n} \binom{n}{i} F_Y^i(y) \left[1 - F_Y(y)\right]^{n-i}, \end{aligned} \tag{32}$$

since the term in the summand is the binomial probability that exactly i of the Y_i are less than or equal to y. If the Y_i are continuous with a density function $f_Y(y)$, the density

function of $\hat{y}_\alpha(n)$ is

$$f_{\hat{y}_\alpha(n)}(y) = \frac{1}{\mathrm{B}(r,n-r+1)} F_Y^{r-1}(y)[1 - F_Y(y)]^{n-r} f_Y(y)$$

where $\mathrm{B}(\cdot,\cdot)$ represents the complete beta function. Unfortunately, while $\hat{y}_\alpha(n)$ is a non-parametric estimator of the α quantile, (32) shows that the distribution of the quantile estimator $\hat{y}_\alpha(n)$ depends not only on n and α but also on the unknown distribution of the underlying Y.

The bias and variance of $\hat{y}_\alpha(n)$ also depend on n, α, and the distribution of the underlying Y. Assume that $F_Y(y)$ is continuous with a density function $f_Y(y)$ which is differentiable and nonzero at y_α. The following result for the expected value of the quantile estimator can be derived from results in David (1970, p. 65):

$$\mathrm{E}[\hat{y}_\alpha(n)] = y_\alpha - \frac{\epsilon}{n f_Y(y_\alpha)} - \frac{\alpha(1-\alpha)}{2(n+2)} \frac{f'_Y(y_\alpha)}{f_Y^2(y_\alpha)} + \mathrm{O}\left(\frac{1}{n^2}\right), \tag{33}$$

where ϵ is a sawtooth function of n and α such that $|\epsilon| < 1$ and $f'(\cdot)$ denotes the derivative of the function $f(\cdot)$. An expansion for the variance of the quantile estimator can be derived in similar fashion as

$$\mathrm{Var}[\hat{y}_\alpha(n)] = \sigma^2_{\hat{y}_\alpha(n)} = \frac{\alpha(1-\alpha)}{(n+2) f_Y^2(y_\alpha)} + O\left(\frac{1}{n^2}\right). \tag{34}$$

The notation $g(n) = \mathrm{O}(1/n^2)$ means the absolute value of $g(n)/(1/n^2)$ remains bounded as n goes to infinity. These asymptotic expansions will be discussed in greater detail in Chapter IV.

There are also well known asymptotic results for $\hat{y}_\alpha(n)$ (Serfling, 1980, sec. 2.3).

- If y_α is the unique solution y of $F(y-) \le \alpha \le F(y)$, then $\hat{y}_\alpha(n) \longrightarrow y_\alpha$ with probability 1 as $n \longrightarrow \infty$.
- If $F_Y(y)$ possesses a density $f_Y(y)$ in a neighborhood of y_α, and $f_Y(y)$ is positive and continuous at y_α, then $\hat{y}_\alpha(n)$ has an asymptotic normal distribution in that

$$F_{\hat{y}_\alpha(n)}(y) \sim \mathrm{N}\left\{ y_\alpha, \left(\frac{\alpha(1-\alpha)}{n f_Y^2(y_\alpha)}\right)^{1/2} \right\} \quad \text{as } n \longrightarrow \infty.$$

- Weiss (1964) proved that under mild conditions, the sample marginal quantiles from a multivariate population with an absolutely continuous joint distribution function have an asymptotic multivariate normal distribution. The asymptotic covariance is a function of the multivariate distribution of the underlying multivariate population. This multivariate result is important because of the role of the joint distribution of the controlled and controlling statistics in the theory of controls for variance reduction.

2. Using Sectioning to Estimate the Variance of a Quantile Estimator

When using the quantile estimator from (31) to calculate a point estimate of the α quantile, one must also estimate the variance or equivalently the standard deviation of the point estimate. One could estimate the density of Y at y_α and use (34) to estimate the variance. However, the instability of density estimates at extreme quantiles can cause this to be a very biased and unstable estimate of the variance of $\hat{y}_\alpha(n)$. A more general technique is to use sectioning to calculate both a point estimate of the quantile and an estimate of the variance of the point estimate. While non-parametric confidence intervals are available for crude quantile estimates (see Mood, Graybill and Boes, 1974, p. 312), the confidence intervals are not appropriate for controlled quantile estimates. A brief discussion of sectioning follows; for a detailed discussion of sectioning see Lewis and Orav (1989, Chap. 9).

Let the random variable $\hat{y}_\alpha(n)$ be the function of independent and identically distributed random variables $Y_1, \ldots, Y_n$ defined in (31) such that $\hat{y}_\alpha(n)$ is a point estimator of y_α. Let $\sigma^2_{\hat{y}_\alpha(n)}$ denote the variance of $\hat{y}_\alpha(n)$. Assume for now that there are a total of $N = m \times n$ independent samples of Y, namely $Y_1, \ldots, Y_n, \ldots, Y_N$. The sectioned point estimator, $\overline{\hat{y}_\alpha}(m,n)$, is constructed as follows:

1. Divide the N samples of the random variable Y into m sections with n samples each where for simplicity $n \times m = N$ (equivalently, replicate a sample of size n, m times).
2. For the jth section, $j = 1, \ldots, m$, use (31) to compute $\hat{y}_{\alpha,j}(n)$.
3. Compute $\overline{\hat{y}_\alpha}(m,n)$ as:

$$\overline{\hat{y}_\alpha}(m,n) = \frac{1}{m}\sum_{j=1}^{m} \hat{y}_{\alpha,j}(n). \tag{35}$$

The point estimator $\overline{\hat{y}_\alpha}(m,n)$ is a sample mean of m independent estimates, each of which is based on n samples.

4. Estimate the variance of $\overline{\hat{y}_\alpha}(m,n)$, namely $\sigma^2_{\overline{\hat{y}_\alpha}(m,n)}$, with the sample variance of the sample mean:

$$S^2_{\overline{\hat{y}_\alpha}(m,n)} = \frac{1}{m(m-1)} \sum_{j=1}^{m} \left\{ \hat{y}_{\alpha,j}(n) - \overline{\hat{y}_\alpha}(m,n) \right\}^2 . \tag{36}$$

One advantage of sectioning to estimate the variance of the quantile estimate over estimating the density is that since the $\hat{y}_{\alpha,j}(n)$ in Step 2 above are i.i.d. and the point estimator $\overline{\hat{y}_\alpha}(m,n)$ is their sample mean, $S^2_{\overline{\hat{y}_\alpha}(m,n)}$ is an unbiased estimate of the variance of the point estimate. Furthermore, if the $\hat{y}_{\alpha,j}(n)$ are approximately normally distributed, one can develop approximate confidence intervals for $\overline{\hat{y}_\alpha}(m,n)$ based on a t-statistic with $m-1$ degrees of freedom. A disadvantage of sectioning is the increase in the bias of the point estimate; the first-order bias predicted by (33) for $\overline{\hat{y}_\alpha}(m,n)$ is m times that for $\hat{y}_\alpha(N)$, a point estimate based on all N samples.

For fixed N, the selection of m and n involves a tradeoff between the bias and the variance of $\overline{\hat{y}_\alpha}(m,n)$ as well as the precision of the estimate of the variance of $\overline{\hat{y}_\alpha}(m,n)$. To minimize the bias in $\overline{\hat{y}_\alpha}(m,n)$, as well as improve the approximation to normality of the individual $\hat{y}_j(n)$, one would like n to be large. A drawback of increasing n is the decrease in precision of the estimate of the variance of the point estimate as well as the decrease in the degrees of freedom, $m-1$, for the t-statistic, which relaxes the confidence interval. Using (34) and (36), one can write the expansion for the variance of the sectioned estimate in terms of m only as

$$\sigma^2_{\overline{\hat{y}_\alpha}(m,n)} = \frac{\sigma^2_{\hat{y}_\alpha(n)}}{m} = \frac{\beta}{(N+2m)} + \frac{m\gamma}{N^2} + O\left(\frac{1}{N^2}\right), \tag{37}$$

where β and γ are constants determined by $F_Y(y)$ and α. The presence of m in both the denominator and the numerator in (37) implies, for fixed N, that the value of m which minimizes the variance is a function of the relative magnitudes of β and γ. If β is small relative to γ, one should choose a small m in order to minimize the variance. The value for m must be at least 2 in order to use (36) to estimate the variance. Values for m and n which will minimize the variance or the mean square error of the point estimate can be determined as functions of terms such as β and γ. However, these terms are in turn functions of the distribution of Y which is unknown. After consideration of the above, Lewis

and Orav (1989, p. 262) suggest as a "rough rule of thumb" to make m between 12 and 20 for samples with N over 1000. This usually gives sufficient precision for the estimate of the variance of $\overline{\hat{y}_\alpha}(m,n)$.

Once m and n have been selected, the variance of the point estimate can be estimated. Equation (34) shows that $\sigma^2_{\hat{y}_\alpha(n)}$ is a decreasing function of n. For fixed m, a decrease in $\sigma^2_{\hat{y}_\alpha(n)}$ will cause a corresponding decrease in $\sigma^2_{\overline{\hat{y}_\alpha}(m,n)}$. A technique for reducing $\sigma^2_{\hat{y}_\alpha(n)}$ without increasing n is the use of linear controls.

C. LINEAR CONTROL OF QUANTILES

1. Single and Multiple Linear Controls

As discussed in Chapter I, linear controls is a variance reduction technique which can be used to reduce the variance of an estimate of a statistic of interest, often a sample mean as in Chapter II. The statistic of interest in this chapter is the quantile estimator $\hat{y}_\alpha(n)$ from (31) and eventually the individual section estimate $\hat{y}_{\alpha,j}(n)$ from (35).

To use a linear control for variance reduction a random variable generated in the simulation, called the control or control variable, which is correlated with $\hat{y}_\alpha(n)$, must be available. The expected value of the control must be known, either exactly or approximately. Let C be a random variable which is generated via simulation. Although an estimator of the α quantile of C is not necessarily the most effective control for a given quantile of Y, for purposes of discussion, the estimator of the α quantile of C as defined in (31), namely $\hat{c}_\alpha(n)$ will be used as the control. The random variable $\hat{c}_\alpha(n)$ is a function of n i.i.d. samples of the random variable C. If $\hat{c}_\alpha(n)$ is generated as part of the simulation that produces the samples of Y it will be called an internal control variable. If $\hat{c}_\alpha(n)$ is generated as output from a different simulation, it will be called an external control variable.

The linear control scheme for variance reduction, with a single quantile estimator as a control, uses the same linear additive combination of the control and its expected value as in Equation (2) of Chapter II to produce a controlled estimate $\hat{y}'_\alpha(n)$. The control function, with coefficient θ, is subtracted off from the uncontrolled or crude estimate $\hat{y}_\alpha(n)$ to produce the controlled estimate as follows:

$$\hat{y}'_\alpha(n) = \hat{y}_\alpha(n) - \theta\{\hat{c}_\alpha(n) - \mathrm{E}[\hat{c}_\alpha(n)]\}. \tag{38}$$

Putting aside the question of sectioning for now, the purpose of using a control is to minimize the variance of the controlled estimate, $\sigma^2_{\hat{y}'_\alpha(n)}$, for a fixed sample size n. If the statistic of interest is $\hat{y}_{\alpha,j}(n)$ from (35), minimizing its variance will, for fixed m, minimize the variance of the section estimate $\overline{\hat{y}_\alpha}(m,n)$. The value of θ which minimizes $\sigma^2_{\hat{y}'_\alpha(n)}$ is still, as in (5) of Chapter II, the regression coefficient from the regression of $\hat{y}_\alpha(n)$ on $\hat{c}_\alpha(n)$;

$$\theta = \frac{\sigma_{\hat{y}_\alpha(n),\hat{c}_\alpha(n)}}{\sigma^2_{\hat{c}_\alpha(n)}} = \frac{\sigma_{\hat{y}_\alpha(n)}}{\sigma_{\hat{c}_\alpha(n)}}\,\rho\left[\hat{y}_\alpha(n),\hat{c}_\alpha(n)\right] \tag{39}$$

where $\sigma_{\hat{y}_\alpha(n),\hat{c}_\alpha(n)}$ is the covariance of $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$ and $\rho[\hat{y}_\alpha(n),\hat{c}_\alpha(n)]$ is the correlation between $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$.

One can use multiple controls for variance reduction where $\hat{c}_\alpha(n)$ and θ become q-dimensional column vectors, $\underline{\hat{c}}_\alpha(n)$ and $\underline{\theta}$ with components $\hat{c}_{\alpha,j}(n)$ and θ_j, for $j = 1,\ldots,q$. With multiple controls for quantile estimators, Equation (38) becomes similar to Equation (1) in Chapter II:

$$\hat{y}'_\alpha(n) = \hat{y}_\alpha(n) - \underline{\theta}^T\left\{\underline{\hat{c}}_\alpha(n) - \mathrm{E}[\underline{\hat{c}}_\alpha(n)]\right\}. \tag{40}$$

2. Use of the Asymptotic Expected Value as an Approximation for the Expected Value of the Control

When using a linear control for variance reduction, the expected value of the control is subtracted from the control variable in the control function as in (38) so that the control function will have a mean of zero. A mean-zero control function is desirable when controlling an unbiased estimator such as a sample mean so that the controlled estimate is also unbiased. However, expected values of quantile estimators are rarely known exactly. If the values of the density function of C and its derivative at c_α are known, the biased expected value of the quantile estimator from (33) can be subtracted in the control function so that the control function does not affect the first-order bias in the controlled quantile estimate. *If the expected value of the biased quantile estimator is not known, it can be approximated by the asymptotic expected value of the estimator;* i.e., the actual quantile value c_α. The value c_α will replace $\mathrm{E}[\hat{c}_\alpha(n)]$ in the control function in (38). While this causes the control function to have order $1/n$ bias, there is already order $1/n$ bias in the

estimate being controlled, $\hat{y}_\alpha(n)$, so that the order of the bias in the controlled estimate is the same as in the uncontrolled estimate.

Even when the biased expected value for the control from (33) is known, it may desirable to use the asymptotic value. There is empirical evidence (to be shown in Section E) that the use of a control function with order $1/n$ bias can actually decrease the magnitude of the first-order bias in the controlled estimate. One can explain this analytically as follows:

1. let $\mathcal{B}_{\hat{y}_\alpha(n)}$ denote the first-order bias of $\hat{y}_\alpha(n)$ computed using (33) as

$$\mathcal{B}_{\hat{y}_\alpha(n)} = \mathrm{E}[\hat{y}_\alpha(n)] - y_\alpha + \mathrm{O}(1/n^2)$$

   and let $\mathcal{B}_{\hat{c}_\alpha(n)}$ denote the bias of $\hat{c}_\alpha(n)$ computed similarly.

2. If using the linear control scheme (38) to control a quantile estimate, let $\mathcal{B}_{\hat{y}'_\alpha(n)}$ denote the first-order bias of $\hat{y}'_\alpha(n)$ so that

$$\mathcal{B}_{\hat{y}'_\alpha(n)} = \mathcal{B}_{\hat{y}_\alpha(n)} - \theta\mathcal{B}_{\hat{c}_\alpha(n)} + \mathrm{O}(1/n^2).$$

The question is then: under what conditions on θ is $|\mathcal{B}_{\hat{y}'_\alpha(n)}| \leq |\mathcal{B}_{\hat{y}_\alpha(n)}|$?

Assume, without loss of generality, that $\mathcal{B}_{\hat{y}_\alpha(n)} > 0$ and that $\mathcal{B}_{\hat{y}_\alpha(n)}/\mathcal{B}_{\hat{c}_\alpha(n)} > 0$. It follows that

$$|\mathcal{B}_{\hat{y}_\alpha(n)'}| \leq |\mathcal{B}_{\hat{y}_\alpha(n)}| \text{ implies that } -\mathcal{B}_{\hat{y}_\alpha(n)} \leq \mathcal{B}_{\hat{y}_\alpha(n)} - \mathcal{B}_{\hat{c}_\alpha(n)} \leq \mathcal{B}_{\hat{y}_\alpha(n)}. \tag{41}$$

Operating on the left-hand side of (41) by dividing both sides of the inequality by $\mathcal{B}_{\hat{y}_\alpha(n)}$, collecting terms and then dividing both sides of the inequality by the ratio $\mathcal{B}_{\hat{c}_\alpha(n)}/\mathcal{B}_{\hat{y}_\alpha(n)}$, one gets that the left side of (41) implies that

$$\theta \leq 2\mathcal{B}_{\hat{y}_\alpha(n)}/\mathcal{B}_{\hat{c}_\alpha(n)}. \tag{42}$$

Operating on the right side of (41) by dividing both sides of the inequality by $\mathcal{B}_{\hat{y}_\alpha(n)}$, collecting terms and then dividing both sides of the inequality by the ratio $\mathcal{B}_{\hat{c}_\alpha(n)}/\mathcal{B}_{\hat{y}_\alpha(n)}$, one gets that the right of (41) implies that

$$0 \leq \theta. \tag{43}$$

Combining (42) and (43), one gets that when $\mathcal{B}_{\hat{y}_\alpha(n)}/\mathcal{B}_{\hat{c}_\alpha(n)}$ is positive and

$$0 < \theta < 2\frac{\mathcal{B}_{\hat{y}_\alpha(n)}}{\mathcal{B}_{\hat{c}_\alpha(n)}},$$

the magnitude of the first-order bias of the controlled estimate is less than the magnitude of the first-order bias of the uncontrolled estimate.

If one is using sectioning to generate the overall point estimate and an estimate of the variance (standard deviation) of the point estimate, and furthermore assumes that θ is known, Equations (35) and (36) can be combined with the linear control equation, (38), to get

$$\overline{\hat{y}'_\alpha}(m,n) = \frac{1}{m}\sum_{j=1}^{m}\hat{y}'_{\alpha,j}(n) \tag{44}$$

$$= \frac{1}{m}\sum_{j=1}^{m}\{\hat{y}_{\alpha,j}(n) - \theta\,(\hat{c}_{\alpha,j}(n) - c_\alpha)\} \tag{45}$$

with an unbiased estimate of the variance of the controlled estimate of

$$S^2_{\overline{\hat{y}'_\alpha}(m,n)} = \frac{1}{m(m-1)}\sum_{j=1}^{m}\left\{\hat{y}'_{\alpha,j}(n) - \overline{\hat{y}'_\alpha}(m,n)\right\}^2. \tag{46}$$

These results are straightforward. It is when θ is not known, the usual case, and has to be estimated using sectioning, that estimating the variance of the controlled estimate requires some care.

3. Estimating the Coefficients

In the usual case in simulation, the values for θ or $\underline{\theta}$ must be estimated since not enough information is known about the joint distribution of $\hat{y}_\alpha(n)$ and $\underline{\hat{c}}_\alpha(n)$ to determine the regression coefficients. For notation's sake, assume that one is using a single control. If using sectioning to estimate the point estimate along with its variance, the sectioned estimates $\hat{y}_j(n)$ and $\hat{c}_j(n)$, for $j = 1,\ldots,m$ are available to use to estimate θ. The regression coefficient θ can be estimated by the least squares regression of $[\hat{y}_{\alpha,j}(n) - \overline{\hat{y}_\alpha}(m,n)]$ on $\theta[\hat{c}_{\alpha,j}(n) - c_\alpha]$ using the regression model

$$[\hat{y}_{\alpha,j}(n) - \overline{\hat{y}_\alpha}(m,n)] = \theta[\hat{c}_{\alpha,j}(n) - c_\alpha] + \epsilon_j, \quad j = 1,\ldots,m \tag{47}$$

where the $\hat{c}_{\alpha,j}(n)$ are considered fixed and ϵ_j is a mean-zero random variable independent of $\hat{c}_{\alpha,j}(n)$. Denote by $\widehat{\theta}(m,n)$ the estimate of θ from a regression which used m estimates for both the dependent variable and the predictor variable, where each of the estimates was based on n independent samples of Y or C as appropriate.

Once $\widehat{\theta}(m,n)$ is computed, the controlled estimate for each section can be computed using (38) as

$$\hat{y}'_{\alpha,j}(n) = \hat{y}_{\alpha,j}(n) - \widehat{\theta}(m,n)\left\{\hat{c}_{\alpha,j}(n) - c_\alpha\right\}. \tag{48}$$

where c_α is the approximation for the expected value of the control. The final controlled section estimate, $\overline{\hat{y}'_\alpha}(m,n)$, can be computed using (44) as the sample mean of the controlled estimates from each section. Unfortunately, estimating the variance of the $\overline{\hat{y}'_\alpha}(m,n)$ with (46) is not as straightforward since the individual $\hat{y}'_{\alpha,j}(n)$ are generally no longer independent because of the common $\widehat{\theta}(m,n)$. The characteristics of the quantile estimates and the variance estimates depend upon the joint distribution of $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$.

a. Subtleties with the Joint Distribution of the Estimators

A key point of linear controls for quantile estimates is that the joint distribution of the statistic being controlled and the control statistic, here $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$, is of primary importance for determining θ and the characteristics of the controlled estimate, *not* the joint distribution of the underlying populations Y and C.

This is in contrast to the use of a linear control for controlling an estimate of the mean, $\bar{y}$, with the sample mean of the control, $\bar{c}$. In this case, one can determine θ as a function of the joint distribution of Y and C since, using (39),

$$\theta = \frac{\mathrm{Cov}[\bar{y},\bar{c}]}{\mathrm{Var}[\bar{c}]} = \frac{\mathrm{Cov}[y,c]}{\mathrm{Var}[c]}.$$

Although the joint distribution of $\bar{y}$ and $\bar{c}$ is different from the joint distribution of Y and C, one can estimate θ using estimates of the population covariances based on the N individual samples. In general, when controlling estimators other than the sample mean, one must estimate the covariances from the joint distribution of the controlled statistic and the control, not the joint distribution of the underlying populations.

b. *Sectioning with the Assumption that the Joint Distribution is Multivariate Normal*

If the joint distribution of $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$ is multivariate normal and $\underline{\theta}$ is estimated, the point estimate of the quantile and the estimate of the variance of the point estimate have several nice properties:

- the controlled estimates for each section, $\hat{y}'_{\alpha,j}(n)$, are i.i.d. since the sample covariance matrix of the $\hat{c}_{\alpha,j}(n)$ is independent of their sample mean.
- $S^2_{\overline{\hat{y}'_\alpha}(m,n)}$, the estimate of the variance of $\overline{\hat{y}'_\alpha}(m,n)$ from (46) where $\hat{y}'_{\alpha,j}(n)$ is computed using (48), is an unbiased estimator, and
- one can develop an unconditional confidence interval for $\overline{\hat{y}'_\alpha}(m,n)$ using the t-statistic following Lavenberg, Moeller and Welch (1982) since conditionally unbiased estimators remain unbiased unconditionally and conditional confidence intervals remain valid unconditionally (see Kendall and Stuart, 1977, p. 379).

When the multivariate normal assumption, or an assumption of spherical symmetry (see Johnson, 1987), is not valid,

- the controlled estimates from each section $\hat{y}'_{\alpha,j}(n)$ are no longer independent since the sample mean and covariance matrix are no longer independent. The controlled estimates also have additional $O(1/m)$ bias from the estimation of θ.
- $S^2_{\overline{\hat{y}'_\alpha}(m,n)}$ from (46) can still be used to estimate the variance of $\overline{\hat{y}'_\alpha}(m,n)$ although it is now biased, and
- even if the $\hat{y}'_{\alpha,j}(n)$ are normally distributed, a confidence interval based on a t-statistic is only approximate because of the lack of independence of the individual section estimates.

One method for maintaining independence between the controlled section estimates at the cost of a loss of variance reduction is to estimate θ independently for each section.

c. *Subsectioning*

An alternative to estimating a single $\widehat{\theta}(m,n)$, which couples the $\hat{y}'_{\alpha,j}(n)$ together so that they are no longer independent, is to generate an individual estimate of θ for each section. This can be done by subsectioning the n samples within the section and calculating quantile estimates within the section to use as data to estimate $\widehat{\theta}_j(v,l)$. More formally, for each jth section, for $j = 1,\ldots,m$,

1. divide the n samples into v subsections of length l where $v \times l = n$, and
2. estimate $\hat{y}_{\alpha,j,k}(l)$ and $\hat{c}_{\alpha,j,k}(l)$ for each kth subsection, for $k = 1,\ldots,v$.

3. Use the v sets of subsection estimates $\hat{y}_{\alpha,j,k}(l)$ and $\hat{c}_{\alpha,j,k}(l)$ from the jth section to estimate $\widehat{\theta}_j(v,l)$ using a regression model similar to (47).

Once $\widehat{\theta}_j(v,l)$ has been estimated, the controlled estimate for the jth section is computed as

$$\hat{y}'_{\alpha,j}(n) = \hat{y}_{\alpha,j} - \widehat{\theta}_j(v,l)\,(\hat{c}_{\alpha,j}(n) - c_\alpha)\,. \tag{49}$$

The equation is similar to (48) only now there is a subscript on $\widehat{\theta}$, which also has different arguments. The final controlled estimate is calculated as before, as a sample mean using (44), and the estimate of variance of the point estimates is calculated using (46).

An advantage of subsectioning is that by using an independent estimate of θ to calculate each section's controlled estimate, the $\hat{y}'_{\alpha,j}(n)$ are now i.i.d.. A disadvantage of using subsectioning is the loss of predicted variance reduction. This occurs for two reasons. The first is that instead of needing one estimate of θ, now m estimates are needed and each additional estimate tends to reduce the achieved percent variance reduction. The second reason is that $\widehat{\theta}(v,l)$ is not an unbiased estimator of the regression coefficient for $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$ since it is calculated using quantile estimates based on l samples. As shown in Part B.1, the distribution of a quantile estimator is a function of the sample size used for the estimate. Thus estimates which are based on $l < n$ samples have a different joint distribution than $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$. There can also be some additional bias in the $\hat{y}'_{\alpha,j}(n)$ from the estimation of θ_j.

d. Splitting and The Jackknife

Other methods which have been used with linear controls for calculating a point estimate and the variance of the point estimate include splitting and the jackknife. Each of these techniques is described in Lewis and Orav (1989, Chap. 9) and in Nelson (1988).

The splitting technique removes the bias caused by estimating θ with the same data being controlled at the cost of reducing the percent variance reduction. Splitting has been described in Tocher (1963, p. 115) and then in Beale (1985). When using sectioning to generate m individual section quantile estimates $\hat{y}_{\alpha,j}(n)$ and $\hat{c}_{\alpha,j}(n)$, for $j = 1,\ldots,m$, the splitting procedure generates an estimate of θ for each section. The estimate of θ for the jth section is computed using all of the section estimates except the jth set of estimates.

The controlled estimate for each section is computed using (49) with $\widehat{\theta}_j(m-1,n)$. The final controlled estimate and its variance are computed as before as the sample mean of the individual controlled section estimates and the sample variance of the sample mean

The splitting estimator eliminates the bias in $\hat{y}'_{\alpha,j}(n)$ due to estimating θ. However, like the sectioning estimator it has the disadvantage that the $\hat{y}'_{\alpha,j}(n)$ are no longer independent. It also has the same disadvantage as the subsection estimator in that m estimates of θ must be computed, reducing the percent variance reduction. The primary purpose for using the splitting estimator has been to eliminate the $O(1/m)$ bias in the controlled estimate from the estimation of θ in non-normal samples when controlling unbiased estimators. Since the quantile estimator already has $O(1/n)$ bias, which is unaffected by splitting, and splitting has no other clear advantages over the section or subsection estimator, splitting was not used.

Lewis and Orav (1989, p.271) describe jackknifing as a method which can remove the $O(1/n)$ bias in $\hat{y}_\alpha(n)$ at the price of uncertainty about the loss of percent variance reduction in small to medium sized samples. For an "m-fold" jackknife estimate, one combines an estimate based on the entire data set, $\hat{y}_{\alpha,0}(N)$, with m estimates, each based on the data set with N/m samples deleted, $\hat{y}_{\alpha,j}(N-m)$, for $j=1,\ldots,m$, to get a set of m 'pseudo values' ${}_{(j)}\hat{y}_\alpha(N-m)$, for $j=1,\ldots,m$. The final jackknife point estimate is the sample mean of the pseudo values. In some circumstances, one can also use the sample variance of the sample mean of the pseudo values as an estimate of the variance of the jackknife point estimate.

The jackknife estimate has an advantage over the section and subsection estimators in that the bias of the quantile estimates is re ed since each pseudo value is based on estimates using $N-m$ instead of N/m samples. Unfortunately it has some disadvantages as well. Lavenberg, Moeller and Welch (1982) examined the use of the jackknife when using a linear control for the sample mean under the assumption of a multivariate normal distribution between the statistic of interest and the control. They found that the jackknifed confidence interval was usually larger and more computationally expensive than the standard linear control based confidence interval. Nelson (1988) compared the performance of several methods for linear control of the mean when the normality assumption was violated and found that the jackknife was usually "dominated" by the splitting estimator.

The jackknife has been used in quantile estimation. Seila (1982) used a 2-fold jackknife for removing the bias of quantile estimates. However he used a sectioning approach, not the jackknife estimate, for estimating the variance of the point estimate. Miller (1974), and Efron and Gong (1983) imply that the jackknife technique may not be an appropriate tool for use with quantile estimation because of the discontinuous, nonlinear nature of quantile estimators such as (31). Our empirical results (presented in the last section of this chapter) confirmed that the jackknife was not suitable for computing quantile estimates and estimates of the variance of the jackknife point estimate because of the high variability of the point estimates and the poor performance of the jackknife estimate of the variance of the jackknife point estimator.

Regardless of the method chosen, estimating the coefficients can result in a reduction in the percent variance reduction as discussed in Chapter II. Like controlling the mean, linear controls for quantile estimates often have low correlation with the quantile of interest. One can use nonlinear controls for quantile estimates in an attempt to improve the effectiveness of the control scheme.

D. NONLINEAR CONTROL OF QUANTILE ESTIMATES

The fundamental definitions of nonlinear controls found in Chapter II carry directly over when Y and C are quantile estimates instead of sample means. Equation (10) becomes

$$\hat{y}'_{\alpha}(n) = \hat{y}_{\alpha}(n) - H(\underline{\hat{c}}_{\alpha}, \underline{\theta}),$$

where $H(\cdot,\cdot)$ is still the linear additive combination of the q transformed quantile estimates $h_j(\hat{c}_{\alpha,j}(n), \underline{\theta})$.

One of the problems in choosing an approximating transformation $h_j(\hat{c}_{\alpha,j}(n), \underline{\theta})$ is that $\mathrm{E}[h_j(\hat{c}_{\alpha,j}(n), \underline{\theta})]$ must be known exactly or approximately. This severely limits the selection of nonlinear transformations available to approximate $h_j^*(\hat{c}_{\alpha,j}(n))$ as the necessary expected values may be intractable to compute or unknown for some transformations. The difficulty in analytically determining the expected value of the transformed control can be greatly reduced when using monotone transformations of quantile estimators as controls. This is the key idea in making the use of nonlinear controls with quantile estimates practical.

1. The Behavior of Quantiles Under Monotone Transformations

Quantiles have a property that is especially useful when working with nonlinear controls. Under strictly monotone transformations of the underlying random variable, the quantiles transform monotonely as well. For example,

- let $h(\cdot)$ be a strictly monotone function with inverse $h^{-1}(\cdot)$,
- let C be a random variable with a continuous, strictly monotone cumulative distribution function such that for all α between zero and one, $F_C^{-1}(\alpha) = c_\alpha$, and
- let $W = h(C)$ be the transformed random variable.

By definition of a quantile,

$$\Pr\{C \leq c_\alpha\} = \alpha \text{ and } \Pr\{W \leq w_\alpha\} = \alpha.$$

Therefore:

$$\begin{aligned} \Pr\{W \leq w_\alpha\} &= \Pr\{h(C) \leq w_\alpha\} \\ &= \Pr\left\{C \leq h^{-1}(w_\alpha)\right\} = \alpha. \end{aligned}$$

This implies that for all α between zero and one,

$$w_\alpha = h(c_\alpha). \tag{50}$$

For example, if C has a Uniform (0,1) distribution with .9 quantile of $c_{.9} = .9$, then the .9 quantile of $W = h(C) = C^2$, namely $w_{.9}$ is equal to $c_{.9}^2 = .9^2 = .81$.

The key point is that the α quantile of a transformed random variable can be found by applying the same transformation to the α quantile of the original random variable.

2. Controlling Quantile Estimates

The fact that quantiles transform monotonely under strictly monotone transformations of the underlying random variable can also be useful in computing the expected value of a transformed quantile estimator. It is important to note that the random variable being transformed is the quantile estimator $\hat{c}_\alpha(n)$ and not the underlying C. For a given nonlinear transformation, it may be possible to compute the expected value of $h(\hat{c}_\alpha(n))$. For example, if C has a Uniform (0,1) distribution, and $h(\hat{c}_\alpha(n))$ is the scaled

power transformation, (Equation (15) from Chapter II, where $\theta = p$ is constrained to be non-negative), then $h(\hat{c}_\alpha(n))$ has a Beta distribution with a known expected value. For other distributions of $\hat{c}_\alpha(n)$, or other transformations $h(\cdot)$, the expected value may not be tractable to compute. This is where the use of strictly monotone transformations can help.

We are interested in the expected value of the transformed quantile estimator. When a strictly monotone transformation is applied to the underlying C, the quantile estimator $\hat{c}_\alpha(n)$ transforms monotonely as well i.e., if $\hat{c}_\alpha(n)$ estimates c_α and $h(C) = W$, with α quantile w_α, then

$$\hat{w}_\alpha(n) = h(\hat{c}_\alpha(n)). \tag{51}$$

From the point of view of the quantile estimator, applying a strictly monotone transformation to a quantile estimator, $\hat{c}_\alpha(n)$ as in (51), yields the same quantile estimate as using the identical transformation on the underlying random variable C and then using (31) to estimate the α quantile. Although for small n

$$\mathrm{E}[h(\hat{c}_\alpha(n))] \neq h(\mathrm{E}[\hat{c}_\alpha(n)]),$$

it is true that as $n \longrightarrow \infty$,

$$\mathrm{E}[h(\hat{c}_\alpha(n))] \longrightarrow h(c_\alpha) \text{ and } h(\mathrm{E}[\hat{c}_\alpha(n)]) \longrightarrow h(c_\alpha)$$

so that asymptotically, the expected value of the transformed quantile estimator is the same as the expected value of the quantile estimator of the transformed underlying random variable.

Since the asymptotic expected values are the same, if the individual transformation functions $h(\cdot)$ in the control function $H(\hat{c}_\alpha(n), \underline{\theta})$ are restricted to strictly monotone transformations, one can approximate $\mathrm{E}[h(\hat{c}_\alpha(n), \underline{\theta})]$ in the nonlinear control function $H(\hat{c}_\alpha(n), \underline{\theta})$, with the asymptotic expected value of the transformed control, namely, the transformed value of the α quantile, $h(c_\alpha, \underline{\theta})$. Calculating $h(c_\alpha, \underline{\theta})$ is trivial since c_α is a constant. Using the asymptotic expected value with the scaled power transformation from Equation (15) of Chapter II, the nonlinear control scheme becomes

$$\hat{y}'_\alpha(n) = \hat{y}_\alpha(n) - \theta_1 \left\{ \frac{\hat{c}_\alpha(n)^{\theta_2} - 1}{\theta_2} - \frac{c_\alpha^{\theta_2} - 1}{\theta_2} \right\}.$$

The use of the approximation introduces bias into the control function, but it is still $O(1/n)$ and may, as in the linear control case, reduce the magnitude of the first-order bias of the controlled estimate. *The key point is that the analytical burden of calculating the expected value of the transformed control has been greatly reduced.* In fact in many cases, this computation would not be possible at all if the transformation were not monotone.

Once the approximating transformations for the $\underline{\hat{c}}_\alpha$ have been selected, one can use either the section or subsection estimator to estimate $\underline{\theta}$ and calculate the final, controlled point estimate $\overline{\hat{y}'_\alpha}(m,n)$ in (44) and an estimate of the variance of the point estimate. Regardless of the method, the coefficients in $\underline{\theta}$ for $h(\underline{\hat{c}}_\alpha, \underline{\theta})$ can be estimated using a nonlinear least-squares regression algorithm as the nonlinear optimizer.

3. Selection of m and n for a Nonlinearly Controlled Section Estimate when $\underline{\theta}$ Must be Estimated

A major factor that must also be considered in the selection of m and n for fixed sample size N is the impact of n, the number of samples used to compute the individual quantile estimates, on the joint normality of the quantile estimates. When computing a controlled section estimate and estimating the coefficients $\underline{\theta}$, the impact of m and n on the variance of the estimate $\underline{\hat{\theta}}(m,n)$ must also be considered.

As previously discussed, given a fixed sample size N the values of m and n which minimize the mean square error of the crude section estimate are a function of the coefficients in the asymptotic expansions for the mean and variance of the estimator, equations (33) and (34). The variance of the controlled estimate $\hat{y}'_\alpha(n)$ is a function of the variance of the estimate of the coefficients $\underline{\theta}$ in addition to the variance of the crude estimate, $\hat{y}_\alpha(n)$, and the variance of the estimate of the control $\hat{c}_\alpha(n)$. In general, the bias and variance of coefficients estimated via least-squares nonlinear regression is a decreasing function of the number of estimates used as data in the regression (see Gallant, 1987, Chap. 1). When using the section estimator, this implies that one would like m, the number of quantile estimates, to be large. However, as m increases for fixed N, n must decrease, increasing the bias and variance of the estimates used as data in the regression. If n is too small, the

bias and variance of the estimates could be such that there is actually very little nonlinear or even linear relationship between the crude and control quantile estimates so that any control scheme is ineffective.

If n, the number of samples in a section, is too large, the joint distribution of the crude and control quantile estimates approaches a joint normal distribution as seen in Part B.1 of this chapter. The impact of the joint normality is that the optimal nonlinear transformation is now the linear transformation of the linear control as seen in Part C.2 of Chapter II and one has lost the increased effectiveness of the nonlinear control. This result is similar to one obtained by Glynn and Whitt (1989) who state that "No improvement in asymptotic efficiency can be achieved by generalizing the notion of control variables from a linear form to a nonlinear setting." They go on to say however, "...this does not preclude the possibility of better performance by nonlinear methods in a small sample context." (Glynn and Whitt, 1989) The key point is that by avoiding the asymptotic joint normality through keeping small the number of samples used to compute the individual quantile estimates, the nonlinear controls can be more effective than the asymptotic linear controls.

When using the subsection estimator, the interplay between m and n changes. One must now consider the impact of choices for v, the number of subsection estimates, and l, the number of samples used to compute a subsection estimate. With the section estimator one wanted the section size m to be large since each estimate is used as a data point in the regression that determines $\underline{\theta}$. For the subsection estimator m is the number of estimates of $\underline{\theta}$ to compute and a large m implies more regression computations that have to be made, as well as a small value for n. For any given value of n, the choice of v and l has slightly different considerations than the choice of m and n for the section estimator. An important consideration for the subsection estimator is that l be "close" to n so that the joint distribution $\hat{y}_\alpha(l)$ and $\hat{c}_\alpha(l)$ will be similar in shape to that of $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$. If the two joint distributions are not similar in shape, then the subsection estimate of $\underline{\theta}$ could be very biased, reducing the effectiveness of the control. This suggests making v as small as possible while still being two to three times the number of coefficients being estimated. If n is too small, the few samples available for the v subsections of length l will force both v and l to be small, resulting in possibly little structure to exploit, or unreliable estimates

of θ, both of which result in ineffective control. The solution would seem to be to make n large.

Making n too large results in the same problems for the subsection estimator as it did for the section estimator. If n is too large, there are few controlled section estimates which reduces the precision of the variance estimate. More importantly, n is still the critical factor for the joint normality of the estimate being controlled and the control estimate. If n is too large, the asymptotic joint normality reduces the effectiveness of the linear control to that of the linear control.

The selection of m and n for a fixed N which minimizes the bias, variance or mean square error of the controlled estimate is a complicated function of many parameters. These parameters include the value of α, the sample size N, and unfortunately, because of the need to estimate $\underline{\theta}$, characteristics of the unknown joint distribution of the underlying populations Y and C. An alternative to attempting to estimate the optimal m and n via a functional approximation is to use graphical methods to assist in the selection of m and n such as in Heidelberger and Lewis (1981). In the experiment described below, for a given fixed sample size N, the results of using different values of n are compared graphically as well as numerically to assist in selecting m and n.

E. THE SIMULATION EXPERIMENT

1. The Factors

A simulation experiment was performed to validate the results in the preceding sections. It used $M = 300$ or $M = 20$ replications to investigate simulation procedures for estimating the α quantile of a distribution and estimating variance of the quantile estimate. The factors in the simulation experiment included the distribution of the underlying population of interest, the value for α, the method of estimating the quantile, the sample size, the choice of m and n for the section estimator and the choice of the m for the m-fold jackknife estimator. All of the computations were performed in the APL2-based statistical computing package GRAFSTAT.

2. The Statistic of Interest

The distribution used in the results presented here was suggested by Hsu and Nelson (1987). The statistic of interest is the estimator for the α quantile of a random

variable Y where

$$Y = \left(\frac{1}{1.01 - X} - \epsilon\right) \div 100$$

and X has a Uniform (0,1) distribution and ϵ has a Uniform (0,.5) distribution and is independent of X. The untransformed control is the estimator of the α quantile of X. The value of α will be .95 for the results presented here. The true value for the .95 quantile of Y, namely $y_{.95}$, is .164167.

Figure 12 shows the nonlinear nature of the relationship between $\hat{y}_\alpha(n)$ and $\hat{x}_\alpha(n)$ for four values of n with the sample size N fixed at 1000. Prior to plotting, the quantile estimates were standardized by subtracting off the sample mean of the quantile estimates from each estimate, and then dividing each estimate by the sample standard deviation of the quantile estimates. Thus the "true" values are zero. The quantile estimates were standardized so that one could visually assess the correlation between the quantile estimator of interest and the control quantile estimator. Note that the scales of the axes in Figure 12 change as n increases to 100, 250 and 500 as the ranges of the standardized quantile estimates become more concentrated about the true values of zero.

For $n = 25$ in Figure 12, the relationship between $\hat{y}_\alpha(n)$ and $\hat{x}_\alpha(n)$ is highly nonlinear. As n increases to 100, 250 and 500 the relationship seems to become more linear as the number of estimates available decreases to just two at $n = 500$ where with only two pairs of estimates, the relationship must appear linear. However, one can see from Figure 13, where $N = 6000$, that even for $n = 1000$ the relationship between $\hat{y}_\alpha(n)$ and $\hat{x}_\alpha(n)$ still has nonlinear tendencies. In all cases, the relationship appears to be one that would be well approximated by a monotone transformation.

3. The Section Estimator versus the Jackknife Estimator

As stated previously, the section estimator was preferred over the jackknife estimator for estimating the α quantile along with an estimate of the variance (standard deviation) of the quantile estimator. Analytically, the section estimator of the variance of the section estimate from (46) is an unbiased estimator and the section estimate of the standard deviation has O(1/m) bias. What follows will show graphically the performance

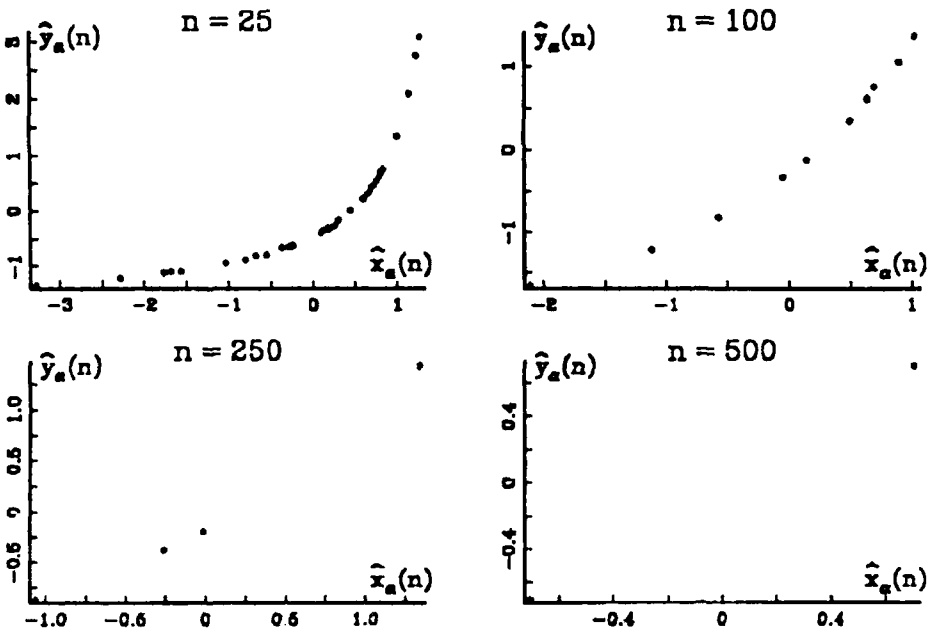


Figure 12. Scatterplots illustrating the joint distribution of standardized section point estimates of the .95 quantile of Y and X for n = 25, 100, 250, and 500 from a sample of N = 1000 samples. Since the estimates are standardized, the true values are zero.

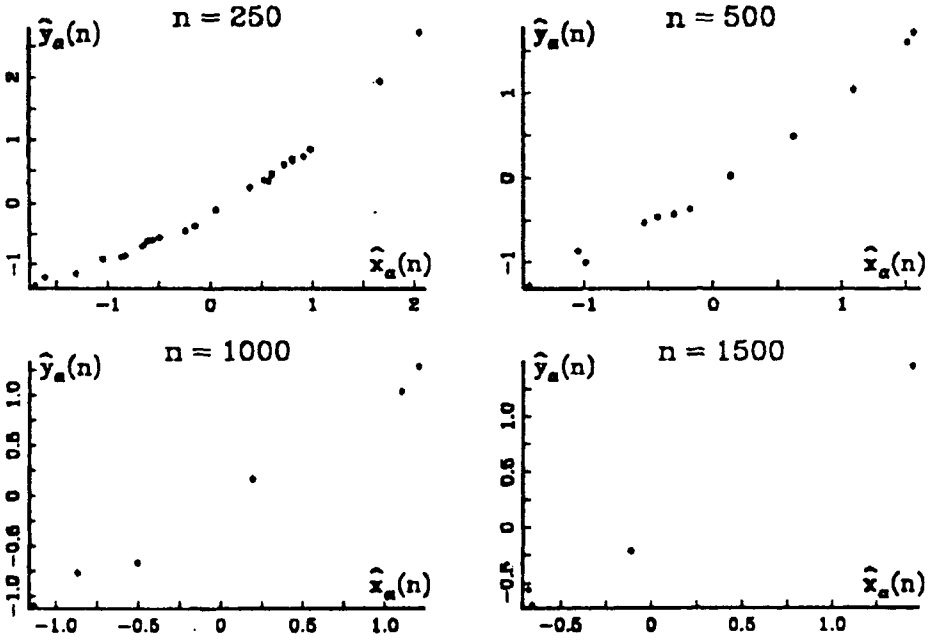


Figure 13. Scatterplots illustrating the joint distribution of of standardized section point estimates of the .95 quantile of Y and X for n = 250, 500, 1000, and 1500 from a sample of N = 6000 samples. Since the estimates are standardized, the true values are zero.

of the section estimate of the standard deviation so that the graphs can be compared with the performance of the jackknife estimation procedure.

One can see the performance of the section estimator in Figure 14. The top graph of Figure 14 shows a series of boxplots of section point estimates of the .95 quantile of Y calculated using (35). (For a discussion of boxplots see Chambers et. al. (1983, Chap. 2).) The boxplots summarize the distribution of the section estimates, for varying n, from 300 independent replications of $N = 1000$ samples. The data under the graph are the sample statistics from the 300 estimates in each boxplot. The bottom graph consists of boxplots of section estimates of the standard deviation, calculated using (36), corresponding to the point estimates in the top graph, again with the sample statistics underneath.

The top graph in Figure 14 shows that as n increases from 10 to 500, for a fixed sample size $N = 1000$, the bias in the section point estimates tends to decrease as expected. However, the top graph also shows that increasing n does not necessarily decrease the sample variance of the section quantile estimator because of the impact of decreasing the number of estimates, m, with which the section point estimate of the quantile is computed.

The bottom graph of Figure 14, of the section estimates of the standard deviation of the section point estimate, shows another effect of increasing n. As n increases and m decreases, it is easy to see that the standard deviation of the estimates of the standard deviation also increases, from .00227 for $n = 10$ to .01170 for $n = 500$, so that the section estimate of the standard deviation becomes less precise. As the section estimate of the standard deviation has $O(1/m)$ bias, one would expect that the section estimate of the standard deviation should be closer to the estimate of the sample standard deviation for small n. A check of the sample standard deviation in the top graph against the mean of the section estimates of the standard deviation in the bottom graph shows that in fact the two values of .02030 and .01974 are fairly close at $n = 10$ and become farther apart as n increases. The significance of the difference will be examined in a moment.

Figure 15 shows the performance of the jackknife estimator for y_α. The top boxplots are the m-fold jackknife estimate of the .95 quantile of Y, for varying m, from the same 300 independent replications of $N = 1000$ samples used for the section estimates in Figure 14. The data under the graph are the sample statistics from the 300 estimates in each boxplot. The bottom graph in Figure 15 consists of boxplots of the corresponding

jackknife estimates of the standard deviation of the jackknife point estimates in the top graph, again with the sample statistics underneath.

The top graph in Figure 15 shows that for a fixed sample size $N = 1000$, the jackknife estimates become highly variable as m increases, as well as having in general a slight positive bias ($y_\alpha = .164167$). The main reason for not using the jackknife technique however is the poor performance of the jackknife estimate of the standard deviation of the point estimate. A check of the sample standard deviation in the top graph against the mean of the jackknife estimates of the standard deviation in the bottom graph shows that the two estimates of the standard deviation become quite far apart as m increases. For $m = 2$ the values are the closest, at .02202 for the sample standard deviation of the point estimate and .01555 for the jackknife estimate of the standard deviation of the point estimate.

The purpose of estimating the standard deviation of the point estimators is to have a measure of the precision of the point estimate. The section and jackknife estimators of the standard deviation of the point estimate are both trying to estimate the standard deviation of a sample of section or jackknife point estimates. To more formally assess their performance the data was used from the 300 independent replications previously shown in Figures 14 and 15. The procedure used for both the section estimates and the jackknife estimates was as follows:

1. The point estimates from the 300 replications were sectioned into 30 independent sections of 10 point estimates each. The sample standard deviation was computed for each of the 30 sections. Thus there were 30 independent estimates of the sample standard deviation for both the section estimates and the jackknife estimates.
2. Likewise, the 300 estimates of the standard deviation were sectioned into 30 independent sections of 10 estimates of the standard deviation each. These 10 standard deviation estimates were averaged to get a single estimate of the standard deviation for each section. Thus there were 30 independent estimates of the standard deviation from the estimator, for both the section estimator and the jackknife estimator.
3. For each of the 30 sections, the mean of the 10 section or jackknife estimates of the standard deviation from step 2 was subtracted from the sample estimate of the standard deviation from step 1 to yield 30 independent estimates of the difference.

If the section or jackknife estimator is a reliable estimate of the sample standard deviation, then the difference of the sample standard deviation and the section or jackknife estimate of the standard deviation should be zero.

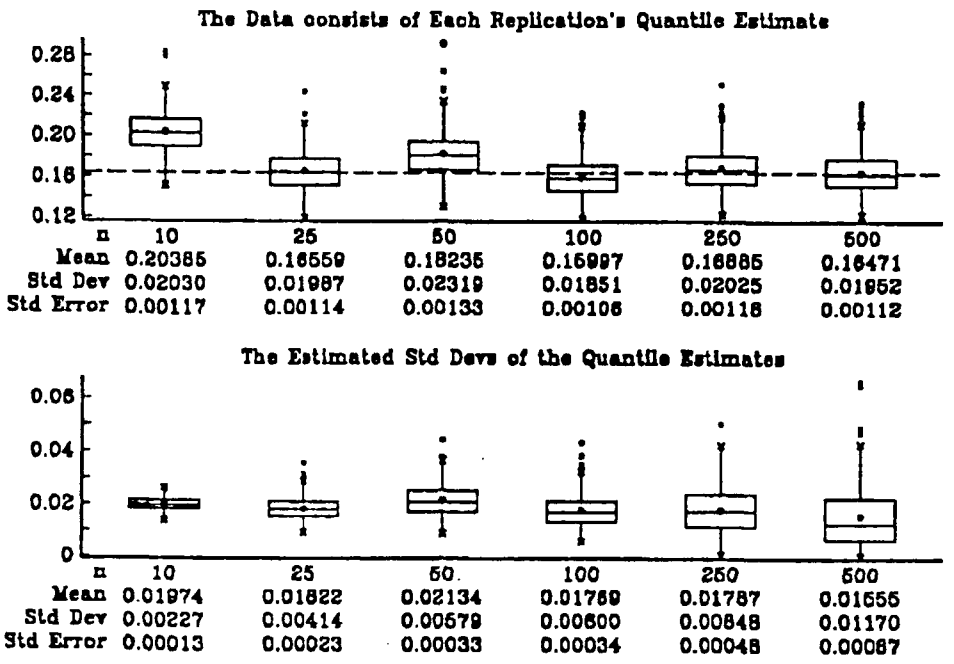


Figure 14. Boxplots of section point estimates of $y_{.95}$ (top) and section estimates of the standard deviation of the point estimates (bottom) for 300 replications of $N = 1000$ samples and varying n.

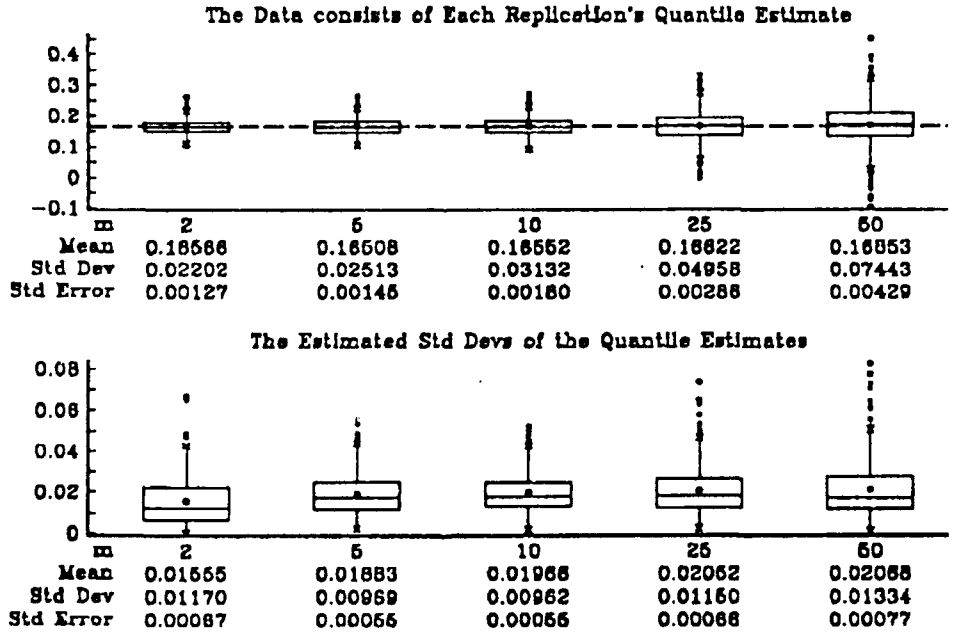


Figure 15. Boxplots of m-fold jackknife point estimates of $y_{.95}$ (top) and m-fold jackknife estimates of the standard deviation of the point estimates (bottom) for 300 replications of $N = 1000$ samples and varying m.

Note that while the same data is used for all of the section and jackknife estimators so that there is no independence between the different estimators, the 30 estimates of the difference for a single estimator i.e., the section estimate with $n = 25$ or the 2-fold jackknife are independent. Figure 16 has boxplots of the differences for both the section estimates (top graph) and the jackknife estimates (bottom graph).

The top graph in Figure 16, of the section estimator, shows that the sample mean for the smaller n is within one standard error of zero. When n is increased to 250 and 500, where the section estimates of the standard deviation are more variable because of the small m, the means of the differences, .00140 and .00300, are still within three standard errors of zero. This shows that section estimator of the standard deviation of the section point estimate is a reliable estimate of the sample standard deviation of the point estimate.

The bottom graph in Figure 16 shows the opposite for the jackknife estimator. For no m is the mean of the differences within three standard errors of zero. If one tests, for each m, the normality of the differences for the jackknife estimates, one can not reject at the .95 confidence level the hypothesis that the differences have a normal distribution. For each m, the .95 confidence interval for the mean of the fitted normal distribution does not include zero. Thus the jackknife estimate of the standard deviation of a jackknifed quantile estimate is a biased and unreliable estimate. This is strong evidence for not using the jackknife technique for estimating quantiles and the variance of the quantile estimate.

4. Comparing the Crude, Linearly Controlled and Nonlinearly Controlled Estimators

The crude, linearly controlled and nonlinearly controlled estimators of y_α will be compared both graphically and numerically. Now the number of replications is $M = 20$ and the number of samples in each replication is fixed at $N = 1000$. The section estimator will be used for all three estimators. For the nonlinearly controlled estimator, the monotone transformation will be the scaled power transformation so that the control function will be

$$\hat{y}'_\alpha(n) = \hat{y}_\alpha(n) - \theta_1 \left\{ \frac{\hat{x}_\alpha(n)^{\theta_2} - 1}{\theta_2} - \frac{x_\alpha^{\theta_2} - 1}{\theta_2} \right\}.$$

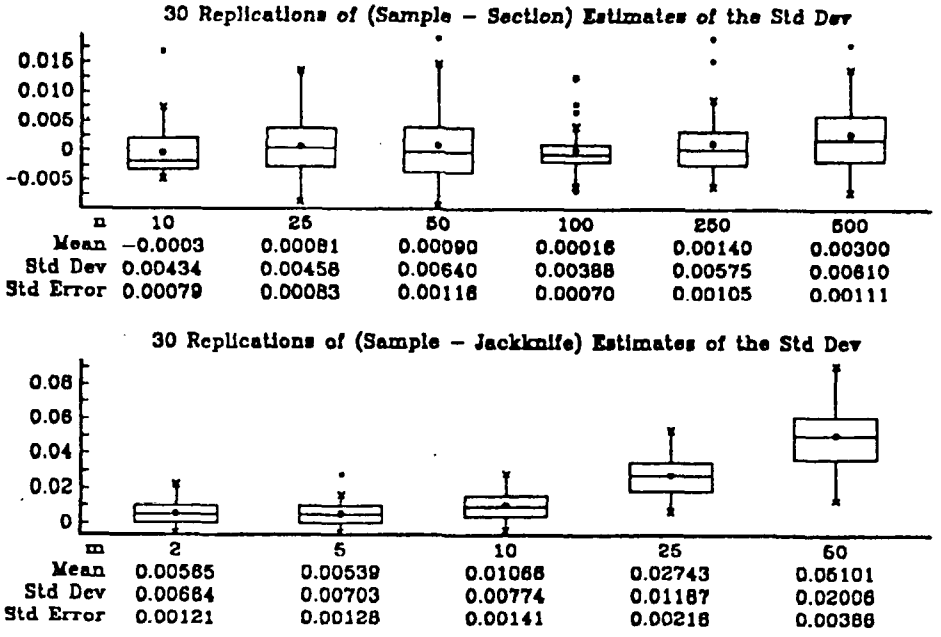


Figure 16. Boxplots of differences between estimates of the sample standard deviation of the point estimate and the section (top) and m-fold jackknife (bottom) estimates of the standard deviation of the point estimate based on 30 sections of $M = 300$ independent replications of $N = 1000$ samples each.

a. *Comparison When the Sample Size $N = 1000$*

Figure 17 shows the performance of the three estimators as triplets of boxplots for $n = 25$, 100, 250, and 500. *In each of the graphs that follow, the left boxplot of the triple is the crude estimate, the middle boxplot of the triple is the linearly controlled estimate and the right boxplot of the triple is the nonlinearly controlled estimate.* The statistics under each graph are the respective means of the data in the boxplot for the crude, linearly controlled and nonlinearly controlled estimators.

The boxplots in the top graph of Figure 17 contain the final quantile estimates for each of the estimators. This graph shows the effect of a control function that is biased because of the use of the asymptotic expected value. Without the biased control function each of the boxplots would look virtually the same because the control function would be mean zero and so would not change the expected value of the point estimate. The

bias in the control function tends to reduce the bias of the point estimate with the exception of the linearly controlled estimate at $n = 25$.

The boxplots in the bottom graph of Figure 17 contain the section estimates of the standard deviation of the point estimators. One can see that as n increases, the mean of the estimated standard deviation of the linearly controlled estimate decreases, from .01123 to .00391, while the mean of the estimated standard deviation for the nonlinear control increases, once n is greater than 100, from .00241 to .00374, until the values for the linear control and the nonlinear control are about the same. In fact, one can see in Figure 17 that the estimator that minimizes the variance is the nonlinearly controlled estimator at $n = 100$ with a value of .00241. It is also clear that when n is large at 250 and 500, the small m of 4 and 2 causes higher variance in the estimates of the standard deviation.

The top graph in Figure 18 combines the two graphs from Figure 17, the bias and the variance, in that it contains the estimated mean square error of the estimators. One can see with this graph that the estimator that minimizes the mean square error is again the nonlinearly controlled estimator at $n = 100$ with a value of .00005. In fact the estimated mean square error for this estimator is under one-half of the best mean square error for the linear control of .00013 that is at $n = 250$. At $n = 500$ the values are the same, .00029, since there are only 2 quantile estimates with which to work. The other factor affecting the nonlinear control besides having only 2 quantile estimates to work with is that at $n = 500$ the joint distribution of the crude estimate and the control estimate is closer to multivariate normal than at $n = 100$.

The bottom graph in Figure 18 is a summary of the percent variance reduction achieved by the various estimators. The percent variance reduction for each estimator is computed using the estimate of the variance of the crude estimate, which is why the value for the crude estimator is 0. This graph again highlights the effectiveness of the nonlinearly controlled estimator at smaller n. The highest percent variance reduction is .97568, which is actually achieved at $n = 25$ and not $n = 100$ because the percent variance reduction is a relative measure and the crude estimator at $n = 25$ had higher variance than the crude estimator at $n = 100$. This graph also points out the high variability of the variance reduction for large n as the number of quantile estimates becomes small.

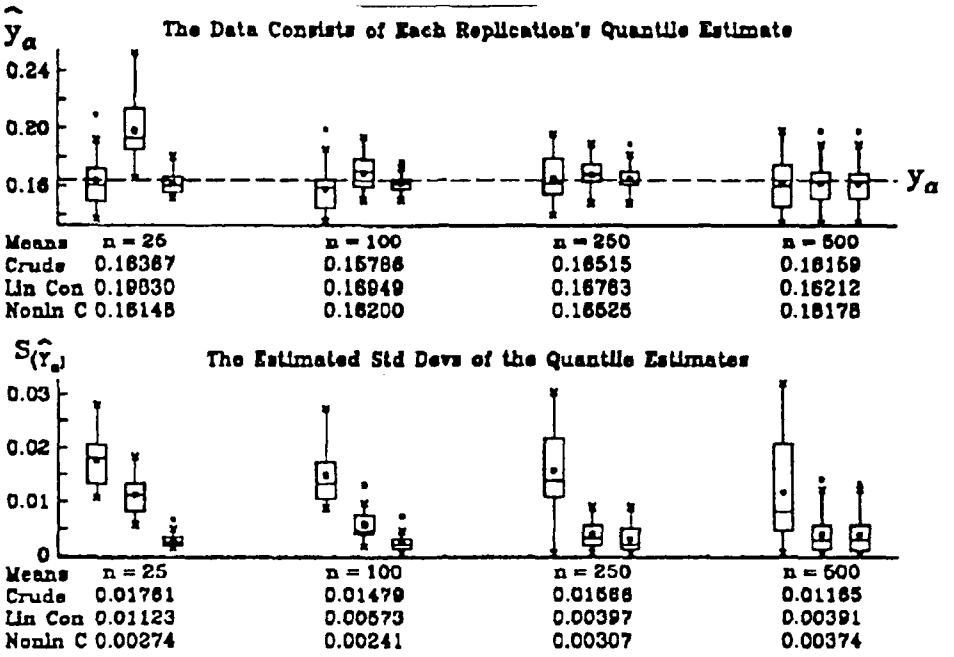


Figure 17. Boxplots of section crude, linearly controlled and nonlinear controlled estimators showing the point quantile estimates of $y_{.95}$ (top) and the estimates of the standard deviation of the point estimates (bottom) from $M = 20$ independent replications of $N = 1000$ for varying n.

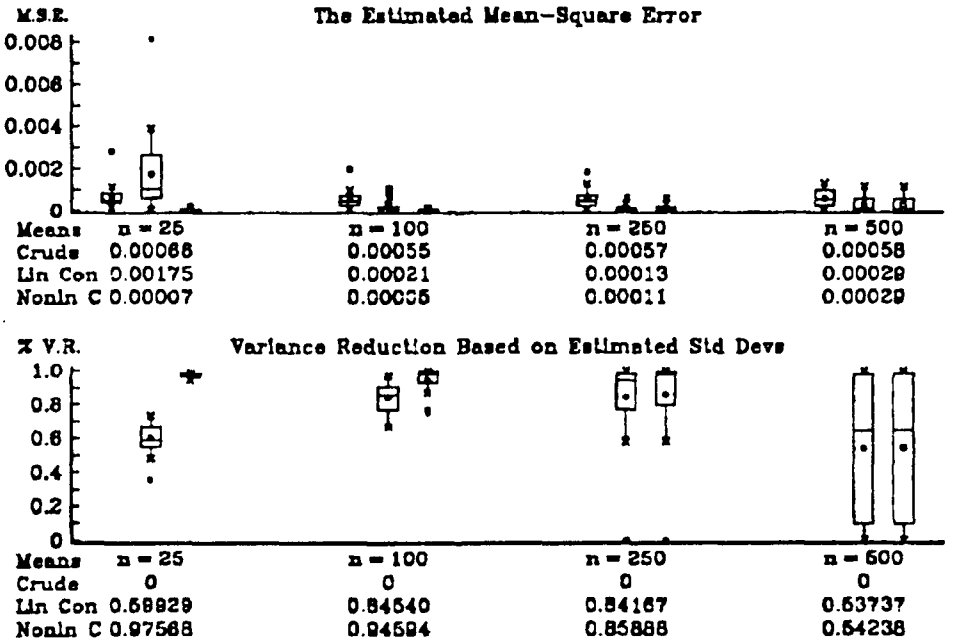


Figure 18. Boxplots of section crude, linearly controlled and nonlinear controlled estimators showing the estimated mean square error (top) and percent variance reduction (bottom) from $M = 20$ independent replications of $N = 1000$ for varying n.

b. *Comparison When the Sample Size $N = 5000$*

The next pairs of graphs, Figures 19 and 20 are identical in nature to the graphs for $N = 1000$ only now the data is from estimates made from a sample size of $N = 5000$. The number of samples used to compute each section estimate n is unchanged so increasing the sample size only increases m, the number of quantile estimates. The larger m greatly reduces the problem of high variability of the estimates caused by having only 2 quantile estimates with which to work at $n = 500$.

One can see by comparing the means of the nonlinearly controlled estimates in the top graph of Figure 17 with those in the top graph of Figure 19 that increasing m has improved the bias of the mean of the nonlinearly controlled estimates. For $n = 25$ in Figure 19 the bias of the mean of the nonlinearly controlled estimate can be computed as .164167 - .16150 =.002. For n= 100, 250, and 500, one gets biases of .001, .004 and .001 respectively. The improvement is such that for each n, the bias of the nonlinearly controlled estimate is now less than the bias of the crude estimate (.004, .003, .008, and .002 respectively). At the same time the bias of the mean of the linearly controlled estimates has increased. A more significant impact of increasing m, shown in the bottom graph, is the drop in the estimated standard deviations for all estimators as compared to $N = 1000$. The variability of the estimates of the standard deviation has decreased as well.

The mean square errors of the top graph in Figure 20 show again that the nonlinear control at $n = 100$ does better than the best linearly controlled estimate. However, as n increases, one can lose the effectiveness of the nonlinear control as both the number of quantile estimates decreases and the quantile estimates approach multivariate normality. One can see the impact of increasing N and m from Figure 18 to Figure 20 in the bottom graph of Figure 20 where the variability of the estimate of the percent variance reduction is greatly reduced as compared to Figure 18.

F. SUMMARY

Nonlinear controls have been seen to be effective in improving the variance reduction over linearly controlled estimates of the mean. Sectioning is a useful procedure for computing point estimates for quantiles along with an estimate of the variance of the point estimate. The jackknife is not a useful procedure as the jackknife estimate of the variance

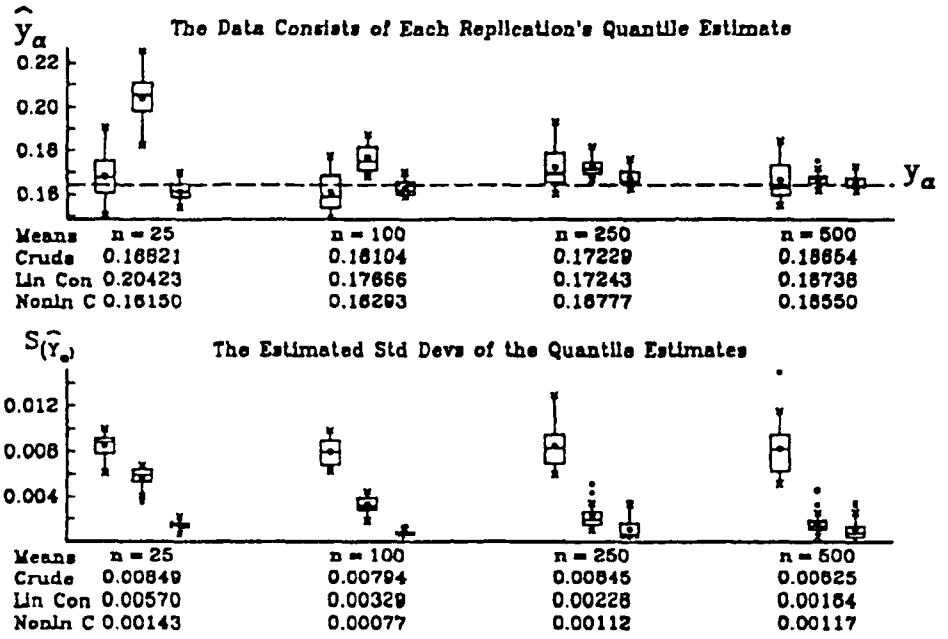


Figure 19. Boxplots of section crude, linearly controlled and nonlinear controlled estimators showing the point quantile estimates of $y_{.95}$ (top) and the estimates of the standard deviation of the point estimates (bottom) from $M = 20$ independent replications of $N = 5000$ for varying n.

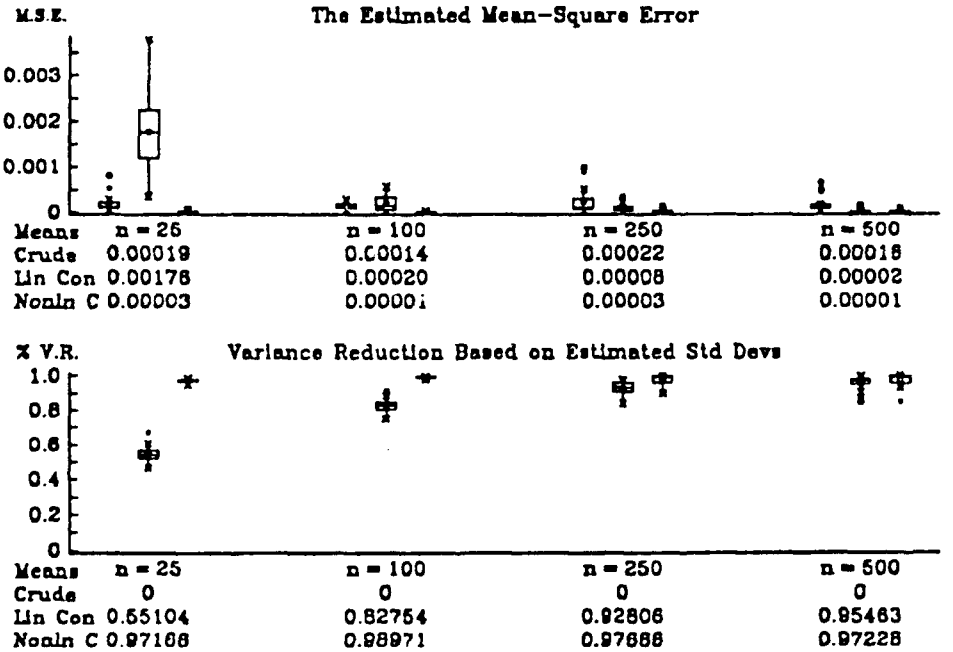


Figure 20. Boxplots of section crude, linearly controlled and nonlinear controlled estimators showing the estimated mean square error (top) and percent variance reduction (bottom) from $M = 20$ independent replications of $N = 5000$ for varying n.

This means that $U_{(r)}$ must be greater than 0 and less than one. Denote the expected value of the rth Uniform order statistic by p_r where

$$p_r = \mathrm{E}\left[U_{(r)}\right] = \frac{r}{n+1}, \quad \text{for } 1 \leq r \leq n. \tag{53}$$

An important property of continuous and strictly monotone functions is described in the following lemma that can be constructed and proved from a problem in Royden (1988, p. 50):

Lemma 1 *Let $h(\cdot)$ be a continuous function on the closed interval $[a,b]$. Then there is a continuous function $l(\cdot)$ such that $l(h(x)) = x$ for all x in $[a,b]$ if and only if $h(\cdot)$ is strictly monotone. It is also true that $h(l(y)) = y$ for all y between $h(a)$ and $h(b)$.*

Note that Lemma 1 implies that $h(\cdot)$ is a bijection (see Definition 4) so that $l(\cdot)$ can be called the inverse of $h(\cdot)$ and written as $h^{-1}(\cdot)$. Also note that $h(\cdot)$ is strictly monotone increasing if and only if $h^{-1}(\cdot)$ is strictly monotone increasing.

Lemma 1 is now used to show that the CDF of X has an inverse.

1. As $\mathcal{D}^*$, the set of x for which $f_X(x)$ is greater than zero, is an open interval, one can choose a closed interval $\mathcal{D} = [a,b]$ where $a < b$ such that $\mathcal{D}$ is a subset of $\mathcal{D}^*$.
2. Since the CDF of X is a continuous and strictly monotone function over $\mathcal{D}^*$, it is also continuous and strictly monotone over $\mathcal{D}$.
3. Thus by Lemma 1, the CDF of X possesses an inverse $F_X^{-1}(u)$ that is continuous and one-to-one from $[F_X(a), F_X(b)]$ to $\mathcal{D}$.

Denote the closed interval $[F_X(a), F_X(b)]$ by $\mathcal{U}$. One can then write the inverse distribution function for X as

$$F_X^{-1}(u) = x \quad \text{for } u \in \mathcal{U}, \tag{54}$$

where since $F_X^{-1}(u)$ is one-to-one, x is an element in $\mathcal{D}$.

Denote $F_X^{-1}(U_{(r)})$ by $Q(U_{(r)})$ so that

$$X_{(r)} = Q\left(U_{(r)}\right), \quad \text{for } U_{(r)} \in \mathcal{U}. \tag{55}$$

Note that for any given value of x with $f_X(x) > 0$, one can choose a and b such that $a < x < b$ and there exists a u such that (54) holds.

One can expand the inverse function $Q(U_{(r)})$ in (55) about the expected value of $U_{(r)}$, namely p_r from (53), in a Taylor series expansion such as in Mood, Graybill and Boes (1974, p. 533). Denote $Q(p_r)$ by Q_r and the first, second, and third derivatives of $Q(U_{(r)})$ with respect to $U_{(r)}$, evaluated at p_r, by Q'_r, Q''_r, Q'''_r and so on. Choosing $\mathcal{D}$ so that it contains p_r, one can then write the Taylor series expansion of $Q(U_{(r)})$ about p_r for (55) as:

$$\begin{aligned} X_{(r)} &= Q_r + \left(U_{(r)} - p_r\right) Q'_r + \frac{1}{2!}\left(U_{(r)} - p_r\right)^2 Q''_r + \frac{1}{3!}\left(U_{(r)} - p_r\right)^3 Q'''_r \\ &\quad + \frac{1}{4!}\left(U_{(r)} - p_r\right)^4 Q''''_r + \cdots, \end{aligned} \tag{56}$$

where

$$\begin{aligned} Q_r &= F_X^{-1}(p_r), \quad \text{for } 0 < p_r < 1, \\ Q'_r &= \frac{1}{f_X(Q_r)} = \frac{1}{f_X(F_X^{-1}(p_r))}, \\ Q''_r &= \frac{-f'_X(Q_r)}{f_X^2(Q_r)}, \\ Q'''_r &= \frac{2f'_X(Q_r)}{f_X^3(Q_r)} - \frac{f''_X(Q_r)}{f_X^2(Q_r)} \quad \text{and} \\ Q''''_r &= \frac{-6f'_X(Q_r)}{f_X^4(Q_r)} + \frac{4f''_X(Q_r)}{f_X^3(Q_r)} - \frac{f'''_X(Q_r)}{f_X^4(Q_r)}. \end{aligned}$$

Note that $f_X(Q_r)$ cannot equal 0 since $F_X^{-1}(p_r)$ is an element of $\mathcal{D}$, which means that $f_X(Q_r)$ is always greater than 0.

The expansion for $X_{(r)}$ in (56) is in terms of powers of $(U_{(r)} - p_r)$. Taking the expected value of $(U_{(r)} - p_r)^m$ yields the mth central moment of the rth Uniform order statistic. Letting $q_r = (1 - p_r)$, one can write the central moments of the rth Uniform order statistic for a given r and n as the constants

$$\mu_1 = \mathrm{E}\left[\left(U_{(r)} - p_r\right)\right] = 0, \tag{57}$$

$$\mu_2 = \mathrm{E}\left[\left(U_{(r)} - p_r\right)^2\right] = \frac{p_r q_r}{n+2}, \tag{58}$$

$$\mu_3 = \mathrm{E}\left[\left(U_{(r)} - p_r\right)^3\right] = \frac{2p_r q_r (p_r - q_r)}{(n+2)(n+3)}, \text{ and} \tag{59}$$

of the jackknife point estimate is unreliable. Controlling quantiles with nonlinear controls is analytically tractable if the nonlinear transformations of the control quantile estimator are limited to strictly monotone functions. With this restriction, one can approximate the expected value of the transformed quantile estimator with its asymptotic expected value, namely the transformed value of the true quantile for the control. The approximation induces additional bias into the control function. However use of a biased control function can reduce the first-order bias in the controlled estimate.

Finally, when one is considering the choice of m and n to use for the sectioning estimator, one must keep n small and avoid approaching the asymptotic multivariate normal distribution. As the joint distribution of the crude estimate of the quantile of interest and the control quantile estimate approaches multivariate normality, the effectiveness of the nonlinear control reduces to that of the linear control.

IV. ASYMPTOTIC EXPANSIONS FOR CONTROLLED QUANTILE ESTIMATORS

A. INTRODUCTION

A key question when controlling one quantile estimator with another quantile estimator is whether the use of a nonlinear control can improve the performance of the control scheme. One can create a nonlinear control by applying nonlinear transformations to the original, linear, control. The question of improved performance is examined in this chapter using asymptotic expansions for the situation where the estimator for the quantile of interest is a strictly monotone transformation of the control quantile estimator. This complements the particular simulation experiment in the last chapter that showed that, for small section sizes, nonlinear controls could improve the variance reduction over linear controls.

In the first sections of this chapter, Sections IV.B through IV.E, definitions are established and asymptotic expansions are constructed for the expected value and the variance of a transformed quantile estimator where the transformations are assumed to be strictly monotone. Sections IV.F and IV.G lead to the construction of an asymptotic expansion for the covariance between two strictly monotone transformations of the same estimator. These expansions are used in section IV.H to construct an expansion for the squared correlation between the two transformed quantile estimators; a direct measure of the effectiveness of one transformed estimator in reducing the variance of the other.

At this point two key results are obtainable. It is shown in section IV.I that given that the statistic of interest is a strictly monotone transformation of the control, by using a nonlinear control instead of a linear control, one can increase the squared correlation between the statistic of interest and the control function thereby improving the variance reduction. It is also shown that asymptotically, as the sample size increases, the nonlinear control loses its advantage over the linear control.

Finally, section IV.J describes a simple example where the expansions predict the improvement gained by the nonlinear control and also give a solution for the optimal parameter for the nonlinear control's transformation. As part of the example, the expansions are

compared against estimates from simulated data. The results from the expansions closely match the results from the simulated data for both the improved performance of the non-linear control over the linear control for small sample sizes and the loss of the improvement as the sample size gets large.

B. DEFINITIONS

1. Definition of an Asymptotic Expansion

Asymptotic expansions can be used to approximate the small sample characteristics of estimators as discussed in Cramér (1966, Chap. 27). Using Taylor series, one can construct asymptotic expansions for functions $h(C)$ of an estimator C from a sample of size n. In what follows, the functions $h(C)$ will usually involve the moments of C i.e., $h(C) = \mathrm{E}[C]$. A general form for an asymptotic expansion of $h(C)$ to order $1/n^m$ for $n \in 1,2,3 \ldots$ and $m \in 1,2,3,\ldots$, is

$$h(C) = h_0 + h_1 n^{-1} + \cdots + h_{m-1} n^{-(m-1)} + \mathrm{O}(\tfrac{1}{n^m}) \tag{52}$$

where $h(C)$ is some function of C such as $\mathrm{E}[C]$ and h_0 through h_{m-1} are constants independent of n. The first few terms usually provide a useful approximation even if the series does not converge for fixed n as the number of terms m gets large (see Barndorff-Neilson and Cox, 1989, Chap. 3). The first term in an expansion, also called the leading term, is the asymptotic value of the function $h(C)$ as the number of samples n gets large.

2. Definitions for Functions

The following definitions, which are taken almost directly from Royden (1988), will be needed in what follows:

Let $f(\cdot)$ be a real-valued function from the set $\mathcal{E}$ to the set $\mathcal{C}$ where the domain $\mathcal{E}$ is a set of real numbers.

Definition 1 *The set of values y in $\mathcal{C}$ for which there exists an x in $\mathcal{E}$ such that $f(x) = y$ is called the* **range** *of $f(\cdot)$ and is denoted by $\mathcal{R}$.*

Definition 2 *If the range of $f(\cdot)$, namely $\mathcal{R}$, is equal to $\mathcal{C}$, then $f(\cdot)$ is a function from $\mathcal{E}$* **onto** *$\mathcal{C}$.*

Definition 3 *A function $f(\cdot)$ from $\mathcal{E}$ to $\mathcal{C}$ is called* **one-to-one** *if $f(x_1) = f(x_2)$ only when $x_1 = x_2$.*

Definition 4 *Functions that are one-to-one and onto are considered* **bijections**. *In this case there is a function $g(\cdot)$ from $\mathcal{R}$ to $\mathcal{E}$ such that for all x in $\mathcal{E}$ and all y in $\mathcal{R}$, it is true that $g(f(x)) = x$ and $f(g(y)) = y$. The function $g(\cdot)$ is called the inverse of $f(\cdot)$ and may be denoted by $f^{-1}(\cdot)$.*

Definition 5 *The function $f(\cdot)$ is* **continuous at the point** *x in $\mathcal{E}$ if given $\epsilon > 0$ there exists a $\delta > 0$ such that for all y in $\mathcal{E}$ with $|y - x| < \epsilon$, it is true that $|f(y) - f(x)| < \delta$.*

Definition 6 *The function $f(\cdot)$ is* **continuous** *on a subset $\mathcal{A}$ of $\mathcal{E}$ if it is continuous at each point of $\mathcal{A}$. Unless stated otherwise, $\mathcal{A}$ will be assumed to be the domain $\mathcal{E}$ of the function $f(\cdot)$.*

Definition 7 *A real-valued function $f(\cdot)$ is* **strictly monotone increasing** *if $f(x) < f(y)$ whenever $x < y$. The function $f(\cdot)$ is called* **strictly monotone decreasing** *if $-f(\cdot)$ is strictly monotone increasing. The function $f(\cdot)$ is called* **strictly monotone** *if either $f(\cdot)$ or $-f(\cdot)$ is strictly monotone increasing.*

C. ASYMPTOTIC EXPANSIONS FOR THE MEAN AND VARIANCE OF A SINGLE ORDER STATISTIC

The results in this section are preliminary and can be found in David (1970) or F.N. David and Johnson (1956).

Let X be a continuous random variable with a cumulative distribution function (CDF) defined by $F_X(x) = \Pr\{X \leq x\}$, for any x that is a real number. Furthermore, let X have a continuous density function $f_X(x) = dF_X(x)/dx$. Let $\mathcal{D}^*$ be the set of x for which $f_X(x)$ is greater than 0. Assume that $f_X(x)$ is such that $\mathcal{D}^*$ is an open interval which is a subset of the real line. Since $f_X(x)$ is continuous, and is positive for every x in $\mathcal{D}^*$, the function $F_x(x)$ is continuous and strictly monotone increasing over $\mathcal{D}^*$.

Let $X_{(r)}$ be the rth order statistic from n independent samples of X. Let $U_{(r)}$ denote the value of the distribution function of X at $X_{(r)}$, namely, $U_{(r)} = F_X(X_{(r)})$. It is straightforward to establish using the probability integral transform that $U_{(r)}$ is distributed as the rth order statistic from a Uniform (0,1) distribution (see David, 1970, p. 16).

$$\mu_4 = \mathrm{E}\left[\left(U_{(r)} - p_r\right)^4\right] = \frac{3p_r^2 q_r^2}{(n+2)^2} + \mathrm{O}\left(\tfrac{1}{n^3}\right) = 3\mu_2^2 + \mathrm{O}\left(\tfrac{1}{n^3}\right). \tag{60}$$

Note that the mth central moment, u_m, has terms of up to order $1/n^{m-1}$.

The expansion in (56) is an infinite sum of powers of random variables and one can integrate each side term-by-term. Taking the expected value of both sides of the expansion of $X_{(r)}$ in (56), one obtains an asymptotic expansion for the expected value of the rth order statistic of X. Using the constants μ_2, μ_3 and μ_4 from (58) through (60), one can write the asymptotic expansion for the expected value of the rth order statistic as

$$\mathrm{E}\left[X_{(r)}\right] = Q_r + \tfrac{1}{2}\mu_2 Q_r'' + \tfrac{1}{6}\mu_3 Q_r''' + \tfrac{1}{24}\mu_4 Q_r'''' + \mathrm{O}\left(\tfrac{1}{n^3}\right). \tag{61}$$

If one knows the density of X and its derivatives, one can compute values for the expansion of the expected value above given r and the sample size n since the values for the central moments of the rth Uniform order statistics are known (David and Johnson, 1956).

By replacing the μ's in (61) with the expressions for the central moments from (58) through (60) and combining terms using the fact that

$$\frac{1}{(n+2)^2} - \frac{1}{(n+2)(n+3)} = \mathrm{O}\left(\tfrac{1}{n^3}\right),$$

one can get the different form of expansion (61) that is in David (1970, p. 65):

$$\mathrm{E}\left[X_{(r)}\right] = Q_r + \frac{p_r q_r}{2(n+2)} Q_r'' + \frac{p_r q_r}{(n+2)^2}\left(\frac{1}{3}(p_r - q_r)Q_r''' + \frac{1}{8}p_r q_r Q_r''''\right) + \mathrm{O}\left(\tfrac{1}{n^3}\right). \tag{62}$$

To construct an expansion for the variance of the rth order statistic, one can use the expression $\mathrm{Var}[X_{(r)}] = \mathrm{E}\left[X_{(r)}^2\right] - \mathrm{E}\left[X_{(r)}\right]^2$. To compute $\mathrm{E}\left[X_{(r)}^2\right]$, one must multiply the Taylor series expansion for $X_{(r)}$ in (56) by itself, prior to taking the expected value, to get an expansion for $X_{(r)}^2$ and then take the expected value of this expansion. Subtracting the square of the expansion for the expected value (61) from the expansion for $\mathrm{E}\left[X_{(r)}^2\right]$, one gets the asymptotic expansion for the variance of $X_{(r)}$,

$$\mathrm{Var}(X_{(r)}) = \mu_2 Q_r'^2 + \mu_3 Q_r' Q_r'' + \mu_2^2\left(Q_r' Q_r''' + \tfrac{1}{2}Q_r''^2\right) + \mathrm{O}\left(\tfrac{1}{n^3}\right). \tag{63}$$

By replacing the μ's with the expressions from (58) through (60) and combining terms in a similar fashion as for the expected value, one can transform (63) into the equivalent expansion for the variance of $X_{(r)}$ found in David (1970, p. 65).

David (1970) and F.N. David and Johnson (1956) cite two cautions when working with these types of expansions. First, when calculating terms for an expansion for the moment of an estimator out to a given order m, which usually involves the $m+1$st central moment of a Uniform random variable, one must also check the expressions for μ_{m+2} and μ_{m+3} for the presence of terms of order m. These higher order moments should be checked to ensure that all terms of order m are actually present in the asymptotic expansion. As an example, note in the expansion for μ_4, (60), that terms of order $1/n^2$ are present in addition to the expected order $1/n^3$ term. Second, these expansions for the moments of order statistics may converge slowly or not at all if n is large and r/n is close to 0 or 1. In general however, the expansions for the moments of order statistics are useful for characterizing the distribution of order statistics and more importantly the distribution of quantile estimators that are based on order statistics.

D. QUANTILE ESTIMATORS AND ORDER STATISTICS

The expansions constructed above are for the central moments of the rth order statistic of a sample. For a fixed sample size n, one can use the expansions to examine the characteristics of a quantile estimator that is based on order statistics. The expansions are not as useful for examining the characteristics of the quantile estimator for changing n because the discontinuous nature of an order statistic based quantile estimator is suppressed in them. The remainder of this section will discuss the order-statistic-based quantile estimator, the implications of making its discontinuous nature explicit in the expansions for the central moments of the quantile estimator, and the reasons for using the expansions for the central moments of the order statistics with the discontinuous nature suppressed in subsequent sections.

Given n independent and identically distributed (i.i.d.) samples of a random variable X, one can estimate the α quantile of X, denoted by x_α, using one of the sample's order statistics. Define the quantile estimator $\hat{x}_\alpha(n)$, as in Equation (31) of Chapter III, as the rth order statistic, where r is either $n\alpha$, if $n\alpha$ is integer, or $(\lfloor n\alpha \rfloor + 1)$ otherwise.

For a fixed n and α, the quantile estimator is fixed as the rth order statistic. Thus the distributions of the quantile estimator and the order statistic are identical as well as the asymptotic expansions for their moments.

However, for changing n, the expansions for the quantile estimator take on a different flavor from the expansions for the order statistic. The asymptotic expansions for the central moments of the rth order statistic in (61) and (63) above are based on Taylor series expansions about the expected value of the rth Uniform order statistic, defined in (53) as $p_r = r/(n+1)$. For the rth order statistic, p_r is a continuous linear function of the sample size n. For a quantile estimator however, the quantity p_r is no longer a linear or even a continuous function of n because of changes in r. For a quantile estimator, one can describe the quantity p_r as a sawtooth function of n that increases with n until $n\alpha$ is integer at which point p_r drops to $n\alpha/(n+1)$ or equivalently to $p_r = \alpha/[1+(1/n)]$. For the rth order statistic, $p_r \to 0$ as $n \to \infty$ while for the quantile estimator $\hat{x}_\alpha(n)$, the quantity $p_r \to \alpha$ as $n \to \infty$.

Since p_r for a quantile estimator is a sawtooth function of n and the asymptotic expansions for the central moments of the rth order statistic are functions of p_r, the expansions for the central moments of the order-statistic-based quantile estimator also exhibit the sawtooth behavior as a function of n when used for computation. One can make the sawtooth behavior an explicit part of the expansions for the central moments of the order-statistic-based quantile estimator through the use of the relationship $\lfloor n\alpha \rfloor + 1 = n\alpha + \epsilon'$ when $n\alpha$ is not integer and where ϵ' is between 0 and 1. To make the sawtooth nature explicit, one must use this relationship to expand the expressions for Q_r, Q_r' and so on in the expansion for the expected value of the rth order statistic (61) as well as expand the expressions for the central moments of the Uniform order statistics in (58) to (60). After making these changes to (61), one gets virtually the same expansion as in Equation (33) of Chapter III, for the expected value of the quantile estimator $\hat{x}_\alpha(n)$, namely

$$\mathrm{E}[\hat{x}_\alpha(n)] = x_\alpha - \frac{\epsilon}{n f_X(x_\alpha)} - \frac{1}{2}\left[\frac{\alpha(1-\alpha)}{n+2} - \frac{\epsilon^2}{n^2}\right]\frac{f_X'(x_\alpha)}{f_X^2(x_\alpha)} + \cdots + \mathrm{O}\left(\tfrac{1}{n^3}\right) \qquad (64)$$

where $|\epsilon| < 1$ is a sawtooth function and the $+\cdots+$ indicates several other terms of order $1/n^2$. One can see in the $1/n$ terms the increase in the notation as compared to (61)

that results from making the sawtooth function explicit, so the other $1/n^2$ terms have been neglected.

Making similar changes to (63), one can derive the expansion in Equation (34) of Chapter III; namely the expansion for the variance of the quantile estimator:

$$\mathrm{Var}\left[\hat{x}_\alpha(n)\right] = \frac{\alpha(1-\alpha)}{(n+2)f_X^2(x_\alpha)} - \frac{(2\alpha-1)\epsilon}{n(n+2)f_X^2(x_\alpha)} + \cdots + \mathrm{O}\left(\tfrac{1}{n^3}\right) \tag{65}$$

where again $|\epsilon| < 1$ is a sawtooth function and the $+\cdots+$ indicates several other terms of order $1/n^2$. Making the sawtooth nature explicit in the expansion for the expected value (64) is useful for seeing that $\hat{x}_\alpha(n)$ is asymptotically unbiased as $n \to \infty$. However, adding the ϵ's makes the both expansions (64) and (65) quite lengthy and cumbersome for further development.

To leave the sawtooth nature of the quantile estimator suppressed in the expansions for the central moments of the order-statistic-based quantile estimator, one can just use the previous definition of r where $r = n\alpha$ or $r = \lfloor n\alpha \rfloor + 1$. Then for changing n p_r is the sawtooth function. Analyzing the performance of the nonlinear control involves fixed sample sizes; it is not an asymptotic issue. For a given α and a fixed sample size n, the expansions for the central moments of the order statistics are appropriate expansions for the central moments of an order-statistic-based quantile estimator. When calculating a value for one of these expansions, n and α must be known so using the suppressed notation with its fewer terms is more expedient and less prone to error. Since the expansions with the sawtooth nature suppressed, (61) and (63), are both notationally and computationally simpler, they will be used for the remainder of the discussion.

E. CONTINUOUS AND STRICTLY MONOTONE TRANSFORMATIONS

In this section, groundwork is laid for the construction of asymptotic expansions for the central moments of quantile estimators (order statistics) of a random variable Y that is a continuous and strictly monotone transformation of the continuous random variable X described in Section C. These expansions will be constructed in terms of the distribution of X that is assumed known. Let $Y = g(X)$ for X in $\mathcal{D}$ where $g(\cdot)$ is a continuous and strictly monotone function over $\mathcal{D}$. Denote the range of $g(X)$ by $\mathcal{R}$. As a continuous and

strictly monotone function over $\mathcal{D}$, the transformation $g(X)$ has an inverse function $g^{-1}(\cdot)$ that is continuous over $\mathcal{R}$ and one-to-one from $\mathcal{R}$ to $\mathcal{D}$ (Lemma 1).

Using the fact that $g^{-1}(\cdot)$ exists, one can express the distribution for Y in terms of the distribution of X and $g^{-1}(\cdot)$. If $g(\cdot)$ is strictly monotone increasing, it follows that (see Mood, Graybill and Boes, 1974, p. 200):

$$F_Y(y) = F_X\left(g^{-1}(y)\right) \quad \text{for } y \in \mathcal{R}, \tag{66}$$

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{dg^{-1}(y)}{dy}\right|, \quad \text{and} \tag{67}$$

$$f'_Y(y) = f'_X\left(g^{-1}(y)\right)\left|\frac{dg^{-1}(y)}{dy}\right| + f_X\left(g^{-1}(y)\right)\left|\frac{d^2g^{-1}(y)}{dy^2}\right|. \tag{68}$$

If $g(\cdot)$ were strictly monotone decreasing, then (66) would become

$$F_Y(y) = 1 - F_X\left(g^{-1}(y)\right), \quad \text{for } y \in \mathcal{R}$$

where $1 - F_X(x)$ is the survivor function (SF) of X. Equations (67) and (68) would change accordingly. Note that $1 - F_X(x)$ is a strictly monotone decreasing function for x in $\mathcal{D}$.

Regardless of the characteristics of $g(\cdot)$, applying the transformation $g(\cdot)$ to n i.i.d. samples of X to get n i.i.d. samples of Y yields n order statistics for Y. However if $g(\cdot)$ is a strictly monotone increasing transformation of X, the transformation maintains the relative order of the sample values so that the order statistics of Y are equal to the transformation $g(\cdot)$ applied to the order statistics of X i.e.,

$$Y_{(r)} = [g(X)]_{(r)} = g(X_{(r)}). \tag{69}$$

If $g(\cdot)$ were strictly monotone decreasing, the transformation $g(X)$ would transpose the relative positions of the order statistics. In this case, equation (69) changes to

$$Y_{(r)} = [g(X)]_{(r)} = g(X_{(\tilde{r})})$$

where $\tilde{r} = \lfloor n(1-\alpha) \rfloor + 1$.

If $g(\cdot)$ is strictly monotone increasing (decreasing), one can consider the distribution function $F_Y(y)$ as the composition of the CDF (SF) of X with the function $g^{-1}(\cdot)$, which

is also strictly monotone increasing (decreasing), where the composition of two functions is defined from Royden (1988, p. 10) as follows:

Definition 8 *If $h_1(\cdot)$ is a function that maps X into Y, and $h_2(\cdot)$ is a function that maps Y into Z, then one can define a new function $l(\cdot)$ that maps X into Z where $l(x) = h_2(h_1(x))$. The function $l(\cdot)$ is called the* **composition** *of $h_2(\cdot)$ with $h_1(\cdot)$ and is denoted by $h_2 \circ h_1(\cdot)$.*

To show that $F_Y(y)$ is continuous and strictly monotone increasing, the following lemma, which can be constructed from a problem in Royden (1988, p. 49), is needed.

Lemma 2 *If $h_1(\cdot)$ and $h_2(\cdot)$ are continuous functions, then $h_2 \circ h_1(\cdot)$ is a continuous function.*

One can now prove the following theorem that shows that as the composition of two continuous and strictly monotone increasing (decreasing) functions, $F_Y(y)$ is also continuous and strictly monotone increasing.

Theorem 1 *If $h_1(\cdot)$ is a continuous and strictly monotone function on $[a,b]$, and $h_2(\cdot)$ is a continuous and strictly monotone function on $[h_1(a), h_1(b)]$, then the function $h_2 \circ h_1(\cdot)$ is continuous and strictly monotone on $[a,b]$.*

The proof shows that as long as $h_1(\cdot)$ or $h_2(\cdot)$ are individually strictly monotone increasing or decreasing, their composition is always strictly monotone.

Proof

1. Since $h_1(\cdot)$ and $h_2(\cdot)$ are both continuous, by Lemma 2 the function $h_2 \circ h_1(\cdot)$ is continuous.
2. Assume $h_1(\cdot)$ is strictly monotone increasing on $[a,b]$ and that $h_2(\cdot)$ is strictly monotone increasing on $[h_1(a), h_1(b)]$. Then

$$x_1 < x_2 \Rightarrow h_1(x_1) < h_1(x_2) \quad \text{for } x_1, x_2 \in [a,b]$$

and

$$h_1(x_1) < h_1(x_2) \Rightarrow h_2(h_1(x_1)) < h_2(h_1(x_2)) \quad \text{for } h_1(x_1), h_1(x_2) \in [h_1(a), h_1(b)].$$

Thus $x_1 < x_2$ implies that $h_2(h_1(x_1)) < h_2(h_1(x_2))$ for all x_1 and x_2 in $[a,b]$, which means that the function $h_2 \circ h_1(\cdot)$ is strictly monotone increasing over $[a,b]$.

3. Assume $h_1(\cdot)$ is strictly monotone increasing on $[a,b]$ and that $h_2(\cdot)$ is strictly monotone decreasing on $[h_1(a), h_1(b)]$. Then using relations similar to those in Step 2, one

establishes that in this case, the function $h_2 \circ h_1(\cdot)$ is strictly monotone decreasing over $[a,b]$.

4. Assume $h_1(\cdot)$ is strictly monotone decreasing on $[a,b]$ and that $h_2(\cdot)$ is strictly monotone increasing on $[h_1(a),h_1(b)]$. Then using relations similar to those in Step 2, one establishes that in this case, the function $h_2 \circ h_1(\cdot)$ is strictly monotone decreasing over $[a,b]$.

5. Assume $h_1(\cdot)$ is strictly monotone decreasing on $[a,b]$ and that $h_2(\cdot)$ is strictly monotone decreasing on $[h_1(a),h_1(b)]$. Then using relations similar to those in Step 2, one establishes that in this case, the function $h_2 \circ h_1(\cdot)$ is strictly monotone increasing over $[a,b]$.

6. In each of the four possible combinations for $h_1(\cdot)$ and $h_2(\cdot)$, the function $h_2 \circ h_1(\cdot)$ is either strictly monotone increasing or strictly monotone decreasing over $[a,b]$. Thus the function $h_2 \circ h_1(\cdot)$ is strictly monotone over $[a,b]$.

■

It is straightforward to show the following corollary to Theorem 1:

Corollary 1.1 *By Lemma 1, since the function $l(\cdot) = h_2 \circ h_1(\cdot)$ is a continuous and strictly monotone function on $[a,b]$, it has an inverse function $l^{-1}(\cdot)$ that is one-to-one such that $l^{-1}[h_2(h_1(x))] = x$ for x in $[a,b]$. It is also true that*

$$l^{-1}(\cdot) = h_1^{-1} \circ h_2^{-1}(\cdot). \tag{70}$$

Since the function $g(\cdot)$ for $Y = g(X)$ is continuous and strictly monotone, by Theorem 1 and Lemma 1, the CDF of the random variable Y has an inverse function $F_Y^{-1}(\cdot)$ that is continuous and one-to-one such that for a given rth order statistic (quantile estimator)

$$Y_{(r)} = F_Y^{-1}(U_{(r)}) = Q_y(U_{(r}$$

where Q_y denotes the inverse of the distribution function for Y.

If $g(\cdot)$ is strictly monotone increasing, it follows from Corollary 1.1 and (70), with $h_1(\cdot) = g^{-1}(\cdot)$ and the CDF (SF) of X as $h_2(\cdot)$, that the inverse function for $F_Y(y)$ is the composition of $g(\cdot)$ with the inverse function for $F_X(x)$ i.e.,

$$Q_y(U_{(r)}) = g\left(F_X^{-1}(U_{(r)})\right) = g \circ Q_x(U_{(r)}) \tag{71}$$

where Q_x denotes the inverse of the distribution function for X. If $g(\cdot)$ is strictly monotone decreasing, then (71) changes to

$$Q_y(U_{(r)}) = Y_{(r)} = g\left(X_{(\tilde{r})}\right) = g\left(F_X^{-1}(U_{(\tilde{r})})\right) = g \circ Q_x(U_{(\tilde{r})})$$

where $\tilde{r} = \lfloor n(1-\alpha) \rfloor + 1$.

The relationships between $Y_{(r)}$ and $X_{(r)}$, shown for $g(\cdot)$ strictly monotone increasing, in (66), (69), and (71), are crucial for what follows. These relationships allow one to develop asymptotic expansions for $Y_{(r)}$ from the known expansions for $X_{(r)}$. The similar relationships for $g(\cdot)$ strictly monotone decreasing also allow one to develop asymptotic expansions for $Y_{(r)}$ from the known expansions for $X_{(\tilde{r})}$. An asymptotic expansion developed for $g(\cdot)$ strictly monotone increasing can be used for $g(\cdot)$ strictly monotone decreasing by the appropriate substitution of the survivor function for the cumulative distribution function and $\tilde{r}$ for r. This generality will be exploited in the following sections by assuming that the function $g(\cdot)$ is strictly monotone increasing and displaying only the one set of expansions.

F. AN ASYMPTOTIC EXPANSION FOR THE MOMENTS OF $Y_{(r)}$ IN TERMS OF THE DISTRIBUTION OF $X_{(r)}$

In this section, asymptotic expansions are constructed for the moments of quantile estimators (order statistics) of the random variable $Y = g(X)$, from the previous section, where $g(\cdot)$ is assumed to be strictly monotone increasing. The expansions are in terms of the known distribution function and moments of X and the central moments of the rth Uniform order statistic.

The random variable $Y_{(r)}$ is an order statistic of a random variable that has a continuous and strictly monotone increasing CDF. As such, one can derive a Taylor series expansion similar to (56) for $Q_y(U_{(r)})$, from the left side of (71), and then expand the expressions such as $Q_y(p_r)$ in terms of $Q_x(p_r) = Q_{xr}$ in order to have the expansion in terms of the distribution of X. As an alternative method, one can consider $g^{-1} \circ Q_x(U_{(r)})$ from the right side of (71) as a composite function and expand it about p_r, using the chain rule to calculate the derivatives.

Let g_r denote the value of $g(Q_{xr})$. Let g_r', g_r'' and so on denote the values of the derivatives of $g(\cdot)$ with respect to $Q_x(U_{(r)})$ evaluated at Q_{xr}. Regardless of the method

used, the asymptotic expansion for $Y_{(r)}$ equivalent to (56) can be written in terms of the distribution of X as

$$\begin{aligned} Y_{(r)} = g(X_{(r)}) &= g_r + (U_{(r)} - p_r)\, g'_r Q'_{xr} + \tfrac{1}{2}\left(U_{(r)} - p_r\right)^2 \left[g'_r Q''_{xr} + g''_r Q'^2_{xr}\right] \\ &\quad + \tfrac{1}{3!}\left(U_{(r)} - p_r\right)^3 \left[g'_r Q'''_{xr} + 2g''_r Q''_{xr} + g'''_r Q'^3_{xr}\right] \\ &\quad + \tfrac{1}{4!}\left(U_{(r)} - p_r\right)^4 \left[g'_r Q''''_{xr} + 3g''_r Q'''_{xr} + 3g'''_r Q''^2_{xr} + g''''_r Q'^4_{xr}\right] + \cdots . \end{aligned} \quad (72)$$

One can now use (72) and the expressions for the central moments from (58) to (60) to compute $\mathrm{E}\left[Y_{(r)}\right]$ and the variance of $Y_{(r)}$ in the same manner as for $X_{(r)}$ in Section C. For example,

$$\begin{aligned} \mathrm{E}\left[Y_{(r)}\right] = \mathrm{E}\left[g(X_{(r)})\right] &= g_r + \tfrac{1}{2}\mu_2 \left[g'_r Q''_r + g''_r Q'^2_r\right] + \tfrac{1}{6}\mu_3 \left[g'_r Q'''_r + 2g''_r Q''_r + g'''_r Q'^3_r\right] \\ &\quad + \tfrac{1}{8}\mu_2^2 \left[g'_r Q''''_r + 3g''_r Q'''_r + 3g'''_r Q''^2_r + g''''_r Q'^4_r\right] + \mathrm{O}\left(\tfrac{1}{n^3}\right), \end{aligned} \quad (73)$$

and to order $1/n^3$,

$$\begin{aligned} \mathrm{Var}\left[Y_{(r)}\right] = \mathrm{Var}\left[g(X_{(r)})\right] &= \mu_2\, g'^2_r Q'^2_r + \mu_3 \left[g'^2_r Q'_r Q''_r + g'_r g''_r Q'^3_r\right] \\ &\quad + \tfrac{1}{2}\mu_2^2 \left[2g'^2_r Q'_r Q'''_r + 6g'_r g''_r Q'^2_r Q''_r + 2g'_r g'''_r Q'^3_r + g'^2_r Q''^2_r + g''^2_r Q'^4_r\right]. \end{aligned} \quad (74)$$

Note that the order $1/n$ and $1/n^2$ factors are contained in μ_2 and μ_3.

The expansion (72) for $Y_{(r)}$ can be used to construct other asymptotic expansions and approximate the moments of any random variable that results from a function $g(X)$ that is continuous and strictly monotone increasing over the X in $\mathcal{D}$. One can use these other expansions for examining the characteristics of the joint distribution between two different transformations of $X_{(r)}$. The joint distribution is of interest as it dictates the effectiveness of a potential control for variance reduction. In the next section, the joint distribution will be examined using an expansion for the covariance between two strictly monotone transformations of $X_{(r)}$.

G. THE COVARIANCE BETWEEN TWO STRICTLY MONOTONE INCREASING TRANSFORMATIONS OF $X_{(r)}$

Let $l(X_{(r)})$ be a function of $X_{(r)}$ that is continuous and strictly monotone increasing over the $X_{(r)}$ in $\mathcal{D}$. Define the covariance between $l(X_{(r)})$ and $g(X_{(r)})$ as

$$\mathrm{Cov}\left[l(X_{(r)}), g(X_{(r)})\right] = \mathrm{E}\left[l(X_{(r)})g(X_{(r)})\right] - \mathrm{E}\left[l(X_{(r)})\right]\mathrm{E}\left[g(X_{(r)})\right].$$

Even though $l(\cdot)$ and $g(\cdot)$ are continuous and strictly monotone transformations of $X_{(r)}$, their product is *not* necessarily a strictly monotone transformation of $X_{(r)}$. Thus one can *not* define a new function $h(X_{(r)}) = l(X_{(r)}) \times g(X_{(r)})$ and use the expansion of (72) with $h(\cdot)$ in place of $g(\cdot)$. If $h(\cdot)$ is not strictly monotone, it does not have an inverse that is one-to-one and continuous; so the expansion does not apply. To calculate $\mathrm{E}\left[l(X_{(r)}) \times g(X_{(r)})\right]$, one must multiply the expansions for $l(X_{(r)})$ and $g(X_{(r)})$, developed individually using (72), and then take the expected value of their product.

To streamline the notation make the following notation changes by dropping the subscript x's and r's from the Q's and X's. Let l denote the value of $l(Q_{xr})$ with its associated derivatives evaluated at Q_{xr} being denoted by l', l'', and so on. Denote the corresponding values for the function $g(\cdot)$ by g, g', g'' and so on. Finally, denote the values and derivatives of the inverse distribution function of X, namely $Q_x(U_{(r)})$, evaluated at p_r by Q, Q', Q'' and so on.

Using the streamlined notation, the expansion for the covariance between two monotone functions of the order statistic has the following form to order $1/n^3$:

$$\mathrm{Cov}\left[l(X_{(r)}), g(X_{(r)})\right] = \mu_2 l'g'Q'^2 + \mu_3\left[l'g'Q'Q'' + \tfrac{1}{2}l'g''Q'^2 + \tfrac{1}{2}l''g'Q'^2\right]$$

$$+\tfrac{1}{2}\mu_2^2\left[2l'g'Q'Q''' + 3l'g''Q'Q'' + 3l''g'Q'Q'' + l'g'''Q'^2 + l'''g'Q'^2 + l'g'Q''^2 + l''g''Q'^2\right]. \qquad (75)$$

Letting $g(X_{(r)}) = l(X_{(r)}) = X_{(r)}$ so that $g' = l' = 1$ and the higher order derivatives are all zero, the expansion above collapses to the expansion for the variance of $X_{(r)}$ in (63) as it should. In the next section the expansion for the covariance will be used to develop an expansion for the squared correlation between two strictly monotone transformations of $X_{(r)}$.

H. AN EXPANSION FOR THE SQUARED CORRELATION BETWEEN TWO STRICTLY MONOTONE TRANSFORMATIONS OF $X_{(r)}$

One can define the correlation between two random variables or estimators $l(X_{(r)})$ and $g(X_{(r)})$ as

$$\rho\left[l(X_{(r)}), g(X_{(r)})\right] = \frac{\mathrm{Cov}\left[l(X_{(r)}), g(X_{(r)})\right]}{\left\{\mathrm{Var}\left[l(X_{(r)})\right]\right\}^{1/2}\left\{\mathrm{Var}\left[l(X_{(r)})\right]\right\}^{1/2}}.$$

The correlation between two estimators determines the effectiveness of one of the estimators in controlling the variance of the other. However, using the squared correlation, an equally valid measure of performance of a control scheme as seen in Equation (6) from Chapter II, avoids the need for radicals in the denominator. Thus one can use the previously developed expansions for the variance of a transformed order statistic (74) and the covariance between two transformations of an order statistic (75) to construct the following expansion for the squared correlation to order $1/n^3$, where the lower order n's are implicit in the μ's:

$$\rho^2\left[l(X_{(r)}), g(X_{(r)})\right] = 1 - \left[\frac{\mu_2}{2} - \frac{\mu_3^2}{4\mu_2^2} - \mu_3\frac{Q''}{Q'}\right]\left[\frac{l''}{l'} - \frac{g''}{g'}\right]^2 + \frac{\mu_3}{2}\left[\frac{l'''(g'-g'')+g'''(l'-l'')}{l'g'}\right]. \tag{76}$$

Several aspects of the expansion above should be noted. First, to order $1/n^2$,

$$\rho^2\left[l(X_{(r)}), g(X_{(r)})\right] = 1 - \frac{\mu_2}{2}\left[\frac{l''}{l'} - \frac{g''}{g'}\right]^2 \tag{77}$$

so that to order $1/n^2$ the squared correlation depends, through the l', l'', g' and g'', only on the value of $Q_{rr} = F_X^{-1}(p_r)$ and not on the shape (derivatives) of the underlying distribution of the random variable X. Second, if $l(\cdot) = g(\cdot)$, the squared correlation is one as is to be expected. Regardless of the relationship between $l(\cdot)$ and $g(\cdot)$ however, the squared correlation is asymptotically one, as all terms in the expansion save the leading term vanish as n increases and the μ's go to zero. This implies that if the quantile estimator being controlled is a strictly monotone transformation of the control, then asymptotically one can achieve complete variance reduction using a linear control or a nonlinear control. However, one can show using the asymptotic expansions (76) and (77) that for small sample sizes, using a nonlinear control can result in a greater squared correlation than a linear control.

The greater squared correlation means that for small sample sizes, the nonlinear control will be more effective at reducing the variance than the linear control.

I. THE RATIO OF SQUARED CORRELATIONS

The squared correlation between the statistic of interest and the control measures the effectiveness of a control scheme for variance reduction. One can use the ratio of the squared correlations to compare the effectiveness of two different controls. In particular, the ratio of the squared correlation for a nonlinear control to the squared correlation for a linear control measures the improvement in variance reduction gained by using a nonlinear control instead of a linear control.

Let the function $l(\cdot)$ be continuous and strictly monotone increasing over the random variable X in $\mathcal{D}$. Let $Y = l(X)$. When estimating the α quantile of Y using the quantile estimator $\hat{y}_\alpha(n)$ defined in Section D, one can use the quantile estimator for the α quantile of X, namely $\hat{x}_\alpha(n)$, as a control. To employ $\hat{x}_\alpha(n)$ as a *linear* control, one uses the identity transformation for $g(\cdot)$ so that $g(\hat{x}_\alpha(n)) = \hat{x}_\alpha(n)$. With this transformation $g' = 1$ and the higher derivatives all equal zero. Now denote the squared correlation between $\hat{y}_\alpha(n)$ and the linear control $g(\hat{x}_\alpha(n))$ as $\rho_L^2[\hat{y}_\alpha(n), \hat{x}_\alpha(n)]$. The asymptotic expansion for this squared correlation simplifies from (76) to

$$\rho_L^2[\hat{y}_\alpha(n), \hat{x}_\alpha(n)] = 1 - \left[\frac{\mu_2}{2} - \frac{\mu_3^2}{4\mu_2^2} - \mu_3\frac{Q''}{Q'}\right]\left(\frac{l''}{l'}\right)^2 + \frac{\mu_3}{2}\frac{l'''}{l'} + O\left(\tfrac{1}{n^3}\right). \tag{78}$$

To use $\hat{x}_\alpha(n)$ as a *nonlinear* control, allow $g(\cdot)$ to be a continuous, strictly monotone increasing transformation of $\hat{x}_\alpha(n)$ other than the identity transformation. Denote by $\rho_N^2[\hat{y}_\alpha(n), g(\hat{x}_\alpha(n))]$ the expansion for the squared correlation from (76) for the nonlinear control. One can now compute the ratio of $\rho_N^2[\hat{y}_\alpha(n), g(\hat{x}_\alpha(n))]$ to $\rho_L^2[\hat{y}_\alpha(n), \hat{x}_\alpha(n)]$ to construct an asymptotic expansion for the improvement in variance reduction gained by using a nonlinear control. One can write this expansion to order $1/n^3$, where the lower order n's i.e., the $1/n$ and $1/n^2$ terms, are implicit in the μ's, as

$$\begin{aligned}\frac{\rho_N^2[\hat{y}_\alpha(n), g(\hat{x}_\alpha(n))]}{\rho_L^2[\hat{y}_\alpha(n), \hat{x}_\alpha(n)]} &= 1 + \left[\frac{\mu_2}{2} - \frac{\mu_3^2}{4\mu_2^2} - \mu_3\frac{Q''}{Q'}\right]\left[\frac{2l''}{l'} - \frac{g''}{g'}\right]\left[\frac{g''}{g'}\right] + \frac{\mu_2^2}{2}\frac{l''g''}{l'g'} \\ &\quad + \frac{\mu_3}{2}\left[\frac{g'''}{g'}\left(1 - \frac{l''}{l'}\right) - \frac{l'''g''}{l'g'}\right].\end{aligned} \tag{79}$$

The nonlinear control will have improved performance over the linear control if the sum of the terms after the 1 in (79) above is positive. Given the number of terms in (79), it is difficult to determine how to choose the transformation $g(\cdot)$ so that the sum of the terms to the right of the 1 is positive. However, one can write the same expansion to order $1/n^2$ as

$$\frac{\rho_N^2\left[\hat{y}_\alpha(n), g(\hat{x}_\alpha(n))\right]}{\rho_L^2\left[\hat{y}_\alpha(n), \hat{x}_\alpha(n)\right]} = 1 + \frac{\mu_2}{2}\left[\frac{2l''}{l'} - \frac{g''}{g'}\right]\left[\frac{g''}{g'}\right]. \tag{80}$$

The constant μ_2 is always positive since it represents the variance of the rth order statistic from a Uniform distribution. Thus by choosing a function $g(\cdot)$ that is continuous and strictly monotone increasing and whose ratio of first and second derivatives evaluated at $F_X^{-1}(p_r)$ are such that the product of the bracketed expressions in (80) is positive, the nonlinear control will have improved performance over the linear control for small samples.

In the standard simulation context one could not use the asymptotic expansions in the straightforward manner described above to select $g(\cdot)$ since very little would be known about the parametric form of the transformation $l(\cdot)$. In addition, the asymptotic expansions do not account for any noise or error that is usually present in simulated data. If the parametric form of $l(\cdot)$ was known i.e., there was no error, and the distribution of X was known, there would be no need for running a simulation. Given that $l(\cdot)$ is not known exactly, one can use other methods for the selection of $g(\cdot)$. One could use graphical analysis to hazard a guess as to the form of $l(\cdot)$ and use the expansions to select $g(\cdot)$. One could also use Breiman and Friedman's ACE algorithm (1985) along with nonlinear least squares regression to assist in the selection of $g(\cdot)$. The key point is that while one can not always use the asymptotic expansions to select $g(\cdot)$, the expansions are useful for demonstrating the potential for improved performance.

In the next section the asymptotic expansions are used in the context where the exact form of $l(\cdot)$ is known. Given parametric forms for $l(\cdot)$ and $g(\cdot)$ the expansions are used to compute the value of the parameter in $g(\cdot)$ that maximizes the improvement in variance reduction. The expansions also predict the improvement one can expect by using this parameter. The expansions are then compared to estimates from simulated data.

J. A SIMPLE EXAMPLE USING THE ASYMPTOTIC EXPANSIONS

The joint distribution of Y and X described in this section is the "noiseless" version of the one in Section E of Chapter III suggested by Hsu and Nelson (1987). The known random variable X has a Uniform (0,1) distribution. The variable of interest Y is equal to the transformation $l(X)$ where

$$Y = l(X) = \frac{1}{(c - X)} \quad \text{for } 0 < X < 1, \tag{81}$$

where c is a constant that is strictly greater than one. Looking at Figure 21, one can see that $l(X)$ is continuous and strictly monotone increasing. Hsu and Nelson (1987) use $c = 1.01$ in (81) as it causes $l(X)$ to be highly nonlinear at the .95 quantile of X. The results that follow also use $c = 1.01$ for the same reason.

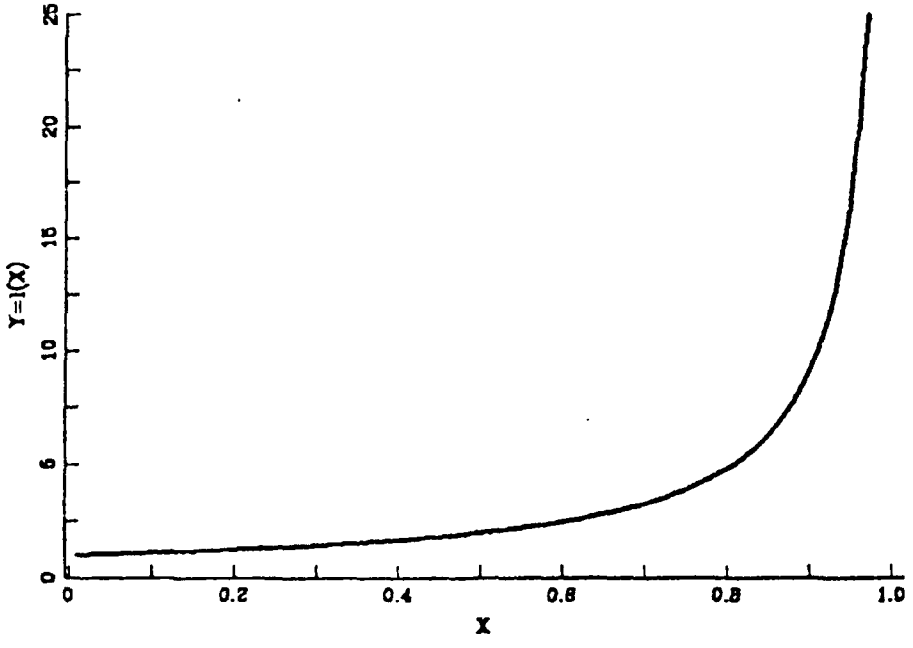


Figure 21. Plot of the transformation $l(X) = 1/(1.01 - X)$.

The quantity to be estimated is the α quantile of Y, namely y_α. It will be estimated using the quantile estimator $\hat{y}_\alpha(n)$ defined in Section D. The control variable is the estimator of the α quantile of X, namely $\hat{x}_\alpha(n)$. In this section the asymptotic expansions will

be used to select the parameter of a transformation $g(\cdot)$ that maximizes the squared correlation between $\hat{y}_\alpha(n)$ and the nonlinear control $g(\hat{x}_\alpha(n))$. The values from the asymptotic expansion will then be compared to simulated data.

1. Specifying the Parameter for $g(\cdot)$

Since X has a Uniform (0,1) distribution, the asymptotic expansions for the variance and covariance (74) and (75) simplify considerably as the only derivative of Q that is nonzero is $Q' = 1$. The expansion for the ratio of the squared correlations to order $1/n^2$, (80), does not change as it does not involve any derivatives of the inverse distribution function.

It follows from (81) that $l'(x) = (1.01 - x)^{-2}$ and $l''(x) = 2(1.01 - x)^{-3}$. Using $\hat{x}_\alpha(n)$ as a linear control is equivalent to using the identity function for $g(\cdot)$, namely $g(\hat{x}_\alpha(n)) = \hat{x}_\alpha(n)$. Thus the derivatives for the linear control are $g'(x) = 1$ and $g''(x) = 0$. Denote by ρ_L^2 the squared correlation between $l(\hat{x}_\alpha(n))$ and its linear control $\hat{x}_\alpha(n)$. The expansion to order $1/n^2$ for ρ_L^2 follows from (78) as simply

$$\rho_L^2 = 1 - [2\mu_2/(1.01 - x)^2] \quad \text{for } 0 < x < 1,$$

where x represents the value for $Q_r = F_X^{-1}(p_r)$, which for the Uniform (0,1) distribution is $p_r = r/(n+1)$.

Now let $g(\cdot)$ be the nonlinear, scaled power transformation, with parameter p,

$$g(x) = (x^p - 1)/p, \quad \text{for } p > -1 \text{ and } 0 < x < 1, \tag{82}$$

with derivatives $g'(x) = x^{p-1}$ and $g''(x) = (p-1)x^{(p-}$ Denote by ρ_N^2 the squared correlation between $Y_{(r)} = l(X_{(r)})$ and its nonlinear control $g(X_{(r)})$. It follows from (77) and (82) that to order $1/n^2$

$$\rho_N^2 = 1 - \frac{\mu_2}{2}\left[\frac{2}{1.01 - x} - \frac{(p-1)}{x}\right]^2 \quad \text{for } p > -1 \text{ and } 0 < x < 1.$$

Let $G(p)$ be the ratio of the squared correlation for the nonlinear control to the squared correlation for the linear control. Thus $G(p)$ is a measure of the relative improvement provided by the nonlinear control over the linear control. Using (80), it

follows that to order $1/n^2$

$$G(p) = 1 + \frac{\mu_2}{2}\left[\frac{4(p-1)}{(1.01-x)x} - \frac{(p-1)^2}{x^2}\right] \quad \text{for } p > -1 \text{ and } 0 < x < 1. \tag{83}$$

One would like to find the value of p for $g(\cdot)$ that maximizes the improvement in variance reduction for a given x; call this value p^*. Differentiating (83), it follows that

$$p^* = \frac{1.01 + x}{1.01 - x} \quad \text{for } 0 < x < 1, \tag{84}$$

since x represents Q_r, which is always between 0 and 1. Substituting p^* into the expansion for $G(p)$ in (83), one gets, to order $1/n^2$, that

$$G(p^*) = 1 + \frac{2\mu_2}{(1.01-x)^2} \geq 1 \quad \text{for } 0 < x < 1, \tag{85}$$

since from (58), $\mu_2 = p_r q_r/(n+2)$ is the variance of the rth Uniform order statistic and is always nonnegative. Thus in this example, for finite samples, the optimal use of a nonlinear control can only improve the variance reduction over the linear control. One can see from (85) that $G(p^*)$ is an increasing function of x so the higher the α value, the greater the improvement gained by using the nonlinear control. However, since μ_2 is a decreasing function of n, as n increases and μ_2 goes asymptotically to zero, the advantage gained by the nonlinear control decreases till $G(p^*)$ is one and the linear and nonlinear controls are equivalent.

2. Comparing the Asymptotic Expansions with Simulated Data

A simulation experiment was used to examine the accuracy of the asymptotic expansions for the example described above. The asymptotic expansions for the variance of $l(x)$ and $g(x)$ from (74), the covariance from (75), and the ratio of squared correlations from (77) were compared to sample estimates of these quantities generated via a simulation experiment.

The two parameters of the simulation were the number of samples used to calculate each quantile estimate, n, and the α value for the quantile estimate. The values for n ranged from 10 to 200. The values for α were .05, .1, .3, .5, .7, .9, .95 and .99. A total of 30 independent replications were used to estimate the precision of each estimate.

Each replication had 4000 i.i.d. Uniform (0,1) random numbers. For each value of α, and each value of n, a total of $4000/n$ quantile estimates were computed. As the maximum value of n in the simulation was 200 there were always at least 20 quantile estimates with which to estimate the variances, covariance and ratio of squared correlations.

For each replication the appropriate transformations, $l(\cdot)$ from (81) with c = 1.01 and $g(\cdot)$ from (82) with $p = p^*$ from (84), were applied to the quantile estimates. The sample variances, covariance, and ratio of squared correlations of the transformed quantile estimates were computed. At the end of the 30 replications, the sample mean and standard error of the 30 estimates were computed for each sample size n and α combination.

Graphs for the function $l(x)$, the function for p^* in (84), and the function $g(x) = (x^p - 1)/p$, Figure 21, Figure 23, and Figure 22, help provide insight into the simulation results. Note in Figure 21 that the function $l(x) = 1/(1.01 - x)$ is fairly flat until $x = .6$ where it curves sharply upward until it is almost vertical at $x = .99$. This shape indicates that the quantile estimators for the lower quantiles of Y will have small variance while those for the upper quantiles will have large variance. Figure 22 shows the function for p^* as a function of x (84). One can see in Figure 22 that as x approaches the higher quantiles, p^* rises dramatically. The reason for this rise can be seen by comparing the plot of $l(x)$ in Figure 21 with the plots of the scaled power transformation for several different vales of p in Figure 23. The curvature of $l(x)$ at the high quantiles forces $g(x)$ to mimic that curvature through the use of higher values of p^*.

One can compare the performance of the asymptotic expansions to estimates from the simulation using Figures (24) through (31). The eight sets of figures, for the eight different values of α, contain four graphs each; the upper left graph showing the variance of $Y = l(X)$ where $l(X)$ is from (81), the upper right graph showing the variance of $g(X)$ where $g(\cdot)$ is the scaled power transformation from (82), the lower left showing the covariance between $l(X)$ and $g(X)$ and the lower right showing the ratio of the squared correlation of $l(X)$ and $g(X)$, the nonlinear control, to the squared correlation of $l(X)$ and X, the linear control. The solid lines on the graphs represent the sample means of the 30 replicates of each statistic. To provide an estimate of the precision of the estimates, dotted lines are plotted three standard errors above and below the sample means. These are often not visible because of the scaling of the graphs. The dashed lines represent the

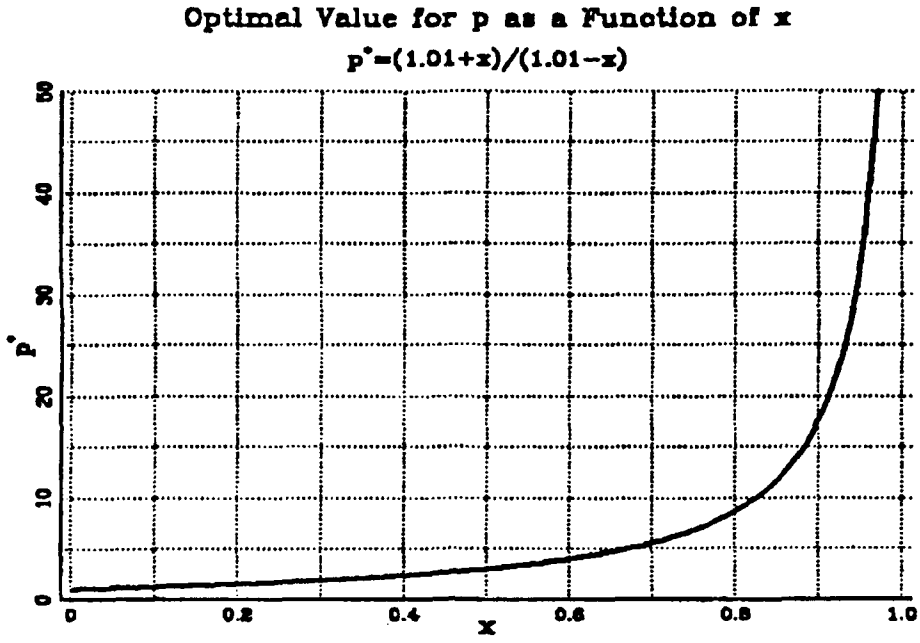


Figure 22. Plot of the optimal value for p, namely p^*, as a function of x for the nonlinear transformation $g(x) = (x^p - 1)/p$.

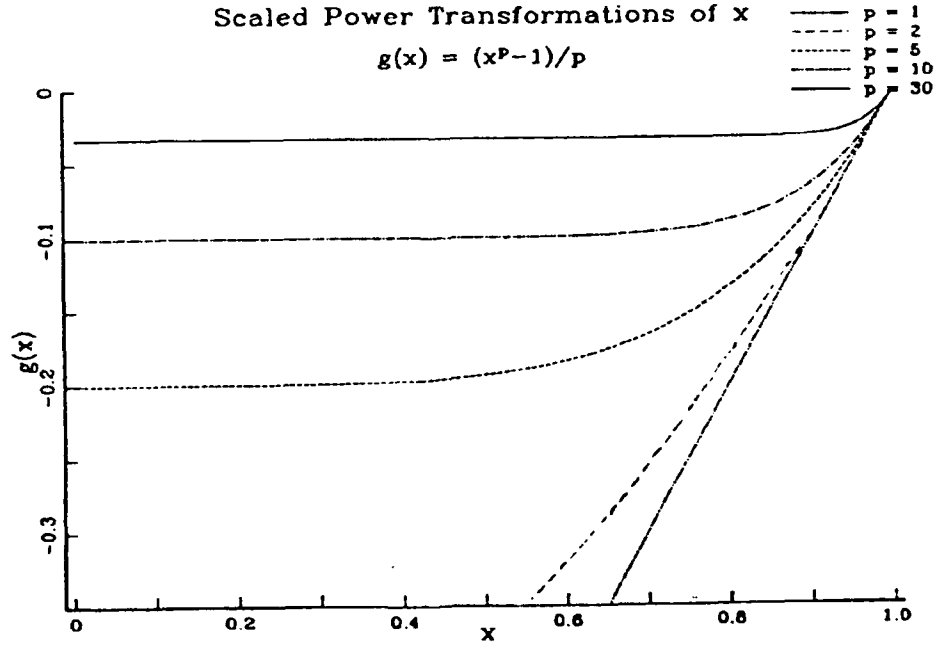


Figure 23. Plot of the nonlinear transformation $g(x) = (x^p - 1)/p$ for several different values of p.

values from the asymptotic expansions. For some of the lower quantiles this dashed line is difficult to see because of the extreme accuracy of the approximations at these quantiles and the scaling of the graph.

In general the figures demonstrate that for this example, the asymptotic expansions for the variance, the covariance and the ratio of squared correlations of the order-statistic-based quantile estimator are quite accurate. The sawtooth nature of the expansions can be seen in most of the figures and is especially noticeable in Figure 30. At the lower quantiles, $\alpha = .05, .1, .3$ and $.5$ in Figures 24 through 27, where the shape of $l(x)$ in Figure 21 is fairly flat, the expansion values for the variance of $l(x)$, the variance of $g(x)$ and the covariance of $l(x)$ and $g(x)$ are accurate even at $n = 10$. While for small n the ratio of squared correlations in Figures 24 through 27 does not match the simulation results as well as the other estimators, by n=50, it matches quite well.

One can see in Figures 24 through 27 that the ratio of squared correlations is a decreasing function of n. Examining the scales of the vertical axes for the ratio of squared correlations in Figures 24 through 27, one can see that the ratio of squared correlations is an increasing function of α; in Figure 24, the upper limit is 1.018 while in Figure 27 the upper limit is 1.18. These values of the ratio of squared correlation indicate that while the nonlinear control does increase the variance reduction over the linear control, the gain is small. The small improvement can be attributed to the fact that $l(x)$ is fairly linear for x between 0 and .6

At the upper quantiles, $\alpha = .7, .9, .95$ and $.99$ in Figures 28 through 31, where the shape of $l(x)$ in Figure 21 is quite nonlinear, the each of the asymptotic expansions show the effects of increasing α. For α =.7 and .9, in Figures 28 and 29, the expansions for the variance and covariance match the estimates from the simulated data for $n > 10$. Comparing the scale of the vertical axes for the graph of the variance of $l(x)$, in the upper left-hand corner of Figures 28 and 29, one can see that the upper limit increases from a variance of 3 to a variance of 80. Comparing the scale of the vertical axes for the graphs of the ratio of the squared correlations, ρ_N^2/ρ_L^2, in the lower right-hand corner of the same figures, the upper limit increases from 1.4 to 2.0. Thus the improvement gained by using a nonlinear control is also increasing with α.

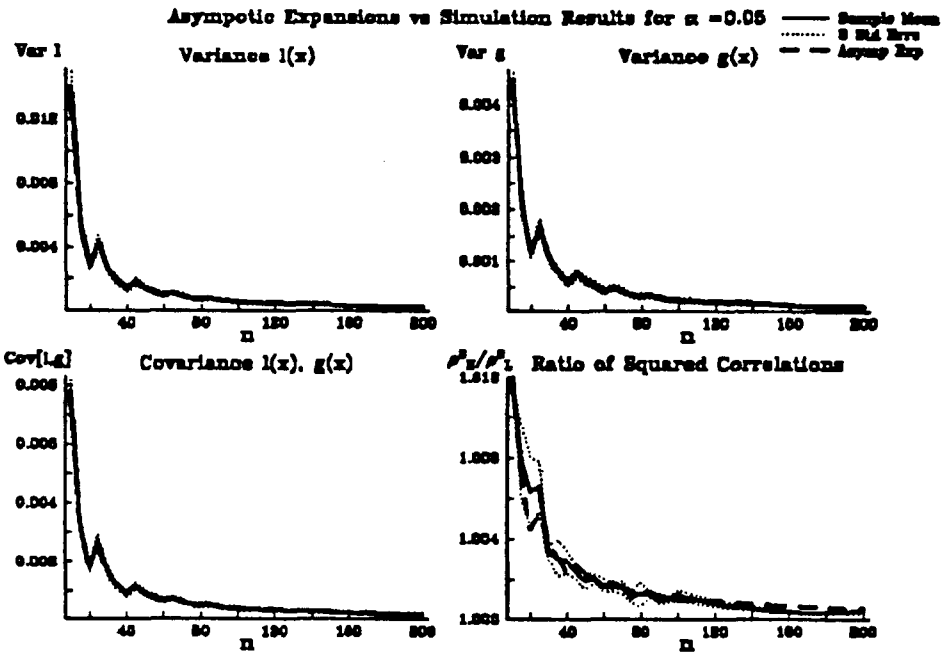


Figure 24. Graphs of the asymptotic expansions for the variance, covariance and ratio of squared correlations compared to estimates from a simulation for the .05 quantile.

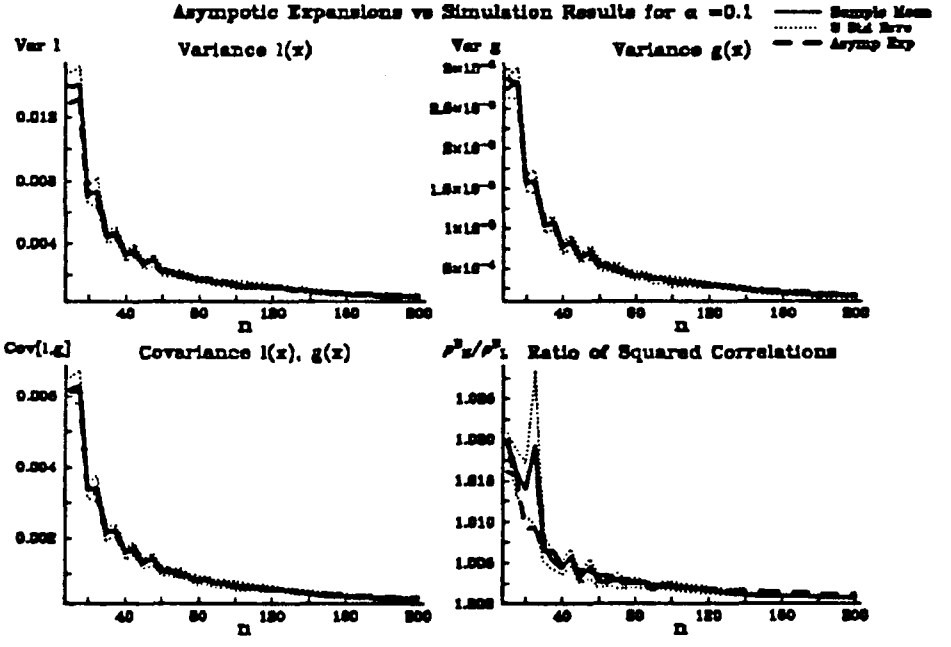


Figure 25. Graphs of the asymptotic expansions for the variance, covariance and ratio of squared correlations compared to estimates from a simulation for the .1 quantile.

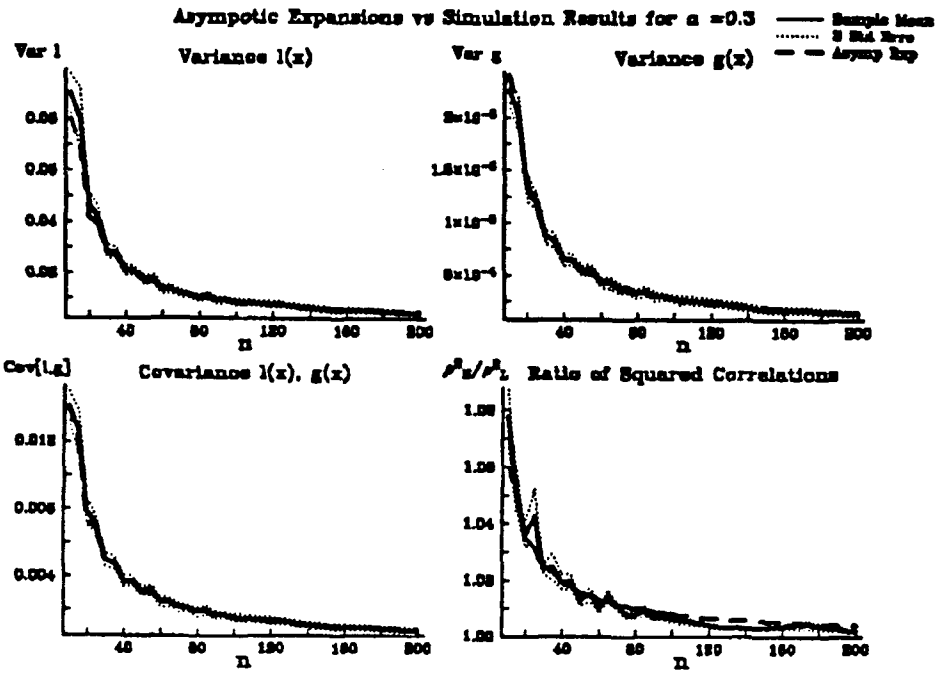


Figure 26. Graphs of the asymptotic expansions for the variance, covariance and ratio of squared correlations compared to estimates from a simulation for the .3 quantile.

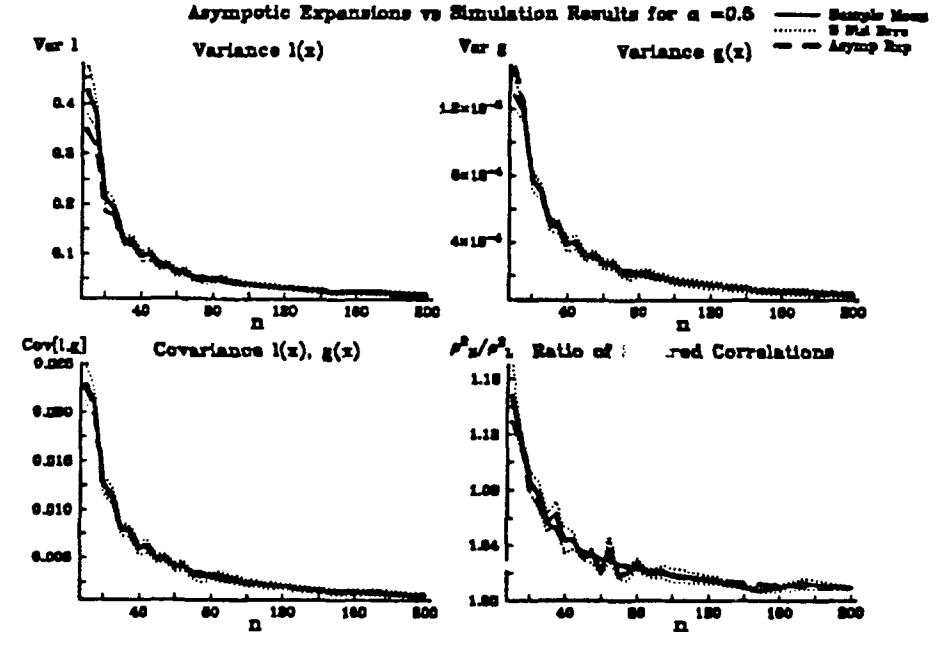


Figure 27. Graphs of the asymptotic expansions for the variance, covariance and ratio of squared correlations compared to estimates from a simulation for the .5 quantile.

At the extreme quantiles, $\alpha = .95$ and .99 in Figures 30 and 31, all three of the asymptotic expansions begin to have difficulty in matching the estimates from the simulation. This difficulty may be due to the nature of the quantile estimator at extreme quantiles. For α=.95, if one is using the quantile estimator defined in Section D with a sample of size less than 20, the order statistic used as the estimator is the maximum of the sample. For $\alpha = .99$ the maximum of the sample is used as the estimator for n less than 100. It is well known that the limiting distribution of the maximum of a sample is different than the limiting distribution of a central or non-extreme order statistic (see Mood, Graybill and Boes, 1974, p. 256). As stated in Section C, David (1970) and David and Johnson (1956) caution against non-convergence of the expansions when r/n is close to 1 for small n. The graphs in Figure 30 for $n < 20$ and in Figure 31 for $n < 100$ demonstrate this weakness in the expansions for the variance, covariance and ratio of the squared correlations for the order-statistic-based quantile estimators.

However, the same graphs also demonstrate that using the value for p^* suggested by the expansions results in an effective nonlinear control for the simulated data. For $\alpha = .95$, in Figure 30, and $\alpha = .99$ in Figure 31, the curve for the ratio of squared correlations shows that at $n = 10$ the nonlinear control is twice as effective as the linear control. While the improvement gained by the nonlinear control decreases as n increases, at $n = 100$ the nonlinear control for $\alpha = .95$ is still 10% more effective than the linear control, and for $\alpha = .99$, the nonlinear control is about 30% percent more effective than the linear control. The variance of $l(x)$ in Figures 30 and 31 doubles, from 400 to 800. At the same time, while the scale of the vertical axes for the ratio of the squared correlations does not change between the two figures, the curve for the mean of the estimates of the ratio of squared correlations from the simulation for $\alpha = .99$ is consistently higher than the curve for $\alpha = .95$. Thus while the expansions do not match the simulation results for small samples at the extreme quantiles, the expansions are useful for choosing the parameter which will create an effective nonlinear control.

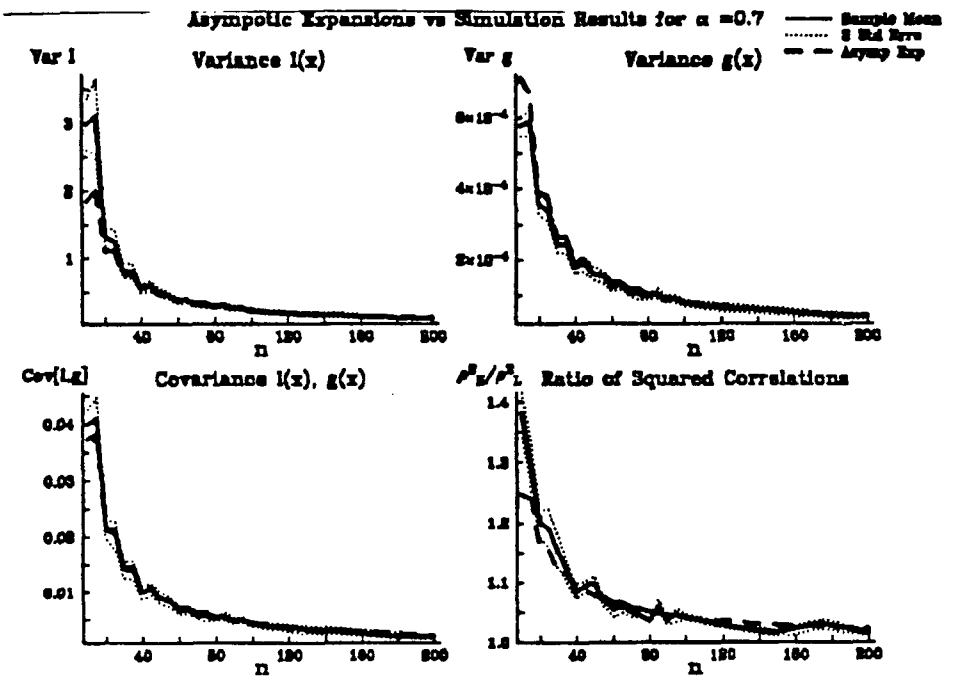


Figure 28. Graphs of the asymptotic expansions for the variance, covariance and ratio of squared correlations compared to estimates from a simulation for the .7 quantile.

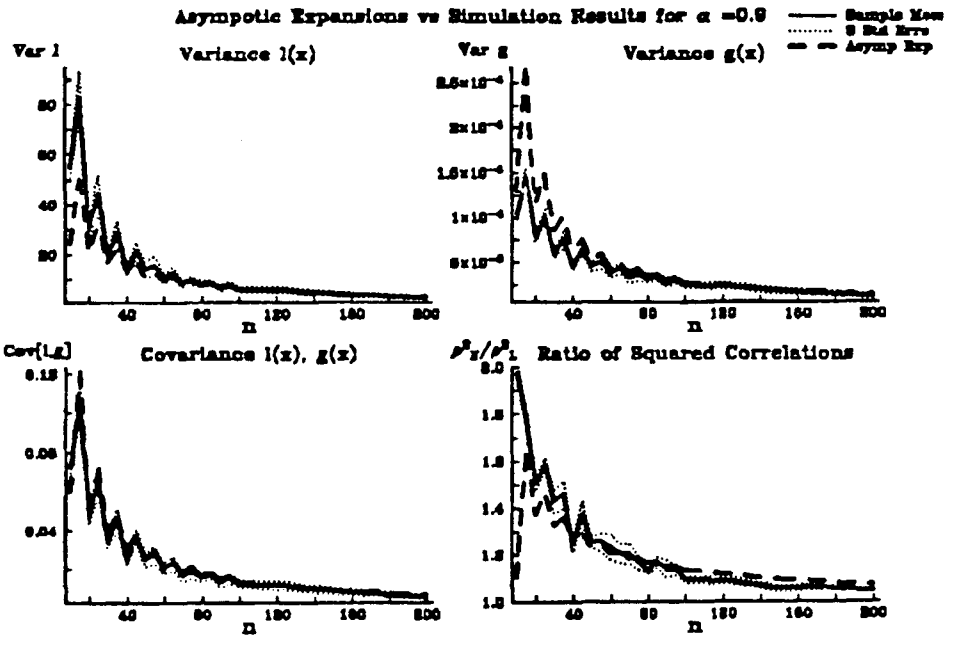


Figure 29. Graphs of the asymptotic expansions for the variance, covariance and ratio of squared correlations compared to estimates from a simulation for the .9 quantile.

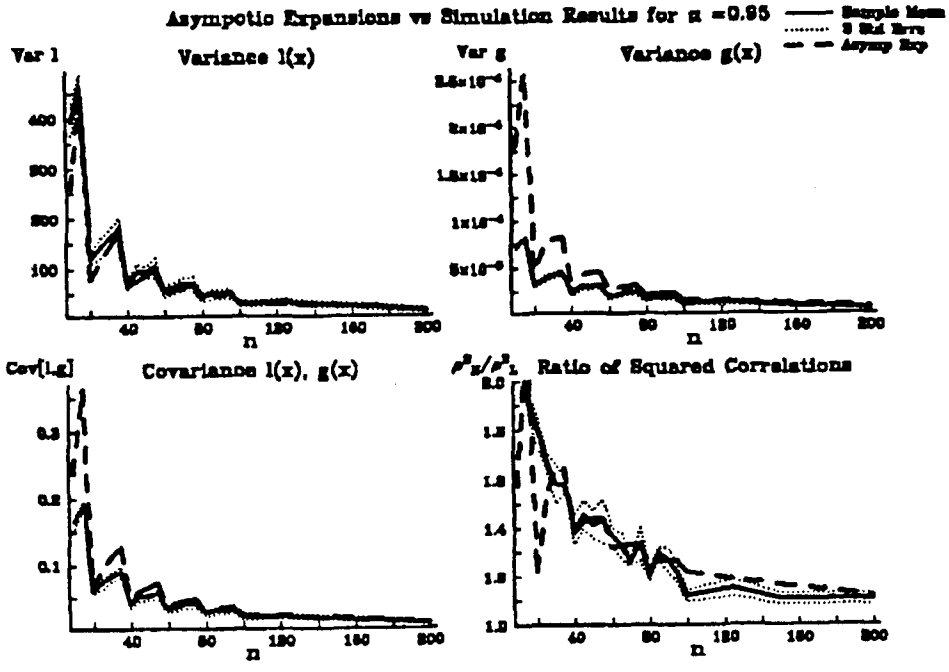


Figure 30. Graphs of the asymptotic expansions for the variance, covariance and ratio of squared correlations compared to estimates from a simulation for the .95 quantile.

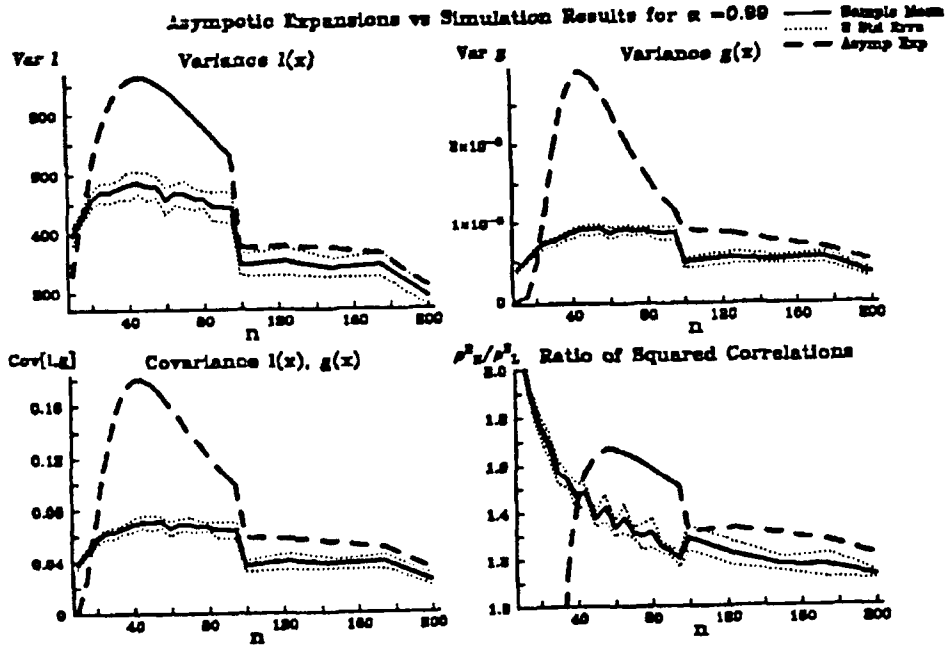


Figure 31. Graphs of the asymptotic expansions for the variance, covariance and ratio of squared correlations compared to estimates from a simulation for the .99 quantile.

K. SUMMARY

The question of the improvement in variance reduction gained by using a nonlinear control instead of a linear control is not an asymptotic issue. The asymptotic joint distribution between the two quantile estimators is bivariate normal so that a linear control scheme is optimal (see Weiss, 1964 and Lancaster, 1966). However, at small sample sizes one can use asymptotic expansions for the moments of order-statistic-based quantile estimators to investigate the relative effectiveness of nonlinear and linear controls. Using the expansions for the variance, covariance and squared correlations constructed in this chapter, one can show that it is possible to choose a transformation such that the squared correlation between the statistic of interest and the nonlinear control is greater than the squared correlation between the statistic of interest and a linear control. This greater squared correlation leads to an improvement in the variance reduction. The comparison of the asymptotic expansions for the variance, covariance and ratio of squared correlations with the estimates from the simulation shows that these expansions can be excellent predictors of the variance and covariance for transformed quantile estimators from a Uniform (0,1) distribution. While for extreme quantiles and very small sample sizes, the expansions may overestimate the improvement gained by the use of a nonlinear control in the simulation, using the parameter indicated by the expansions for the nonlinear control did lead to an improvement in variance reduction for all quantiles. Given that the statistic of interest is a strictly monotone transformation of the control, the asymptotic expansions in this chapter demonstrate that for small samples, the use of a nonlinear control for controlling quantile estimates in a simulation can increase the variance reduction over using a linear control.

V. REGENERATIVE SYSTEM SIMULATION: NONLINEAR CONTROLS AND REGRESSION-ADJUSTED REGENERATIVE ESTIMATES

A. THE CONTROL OF REGENERATIVE ESTIMATES FOR VARIANCE REDUCTION

When simulating queueing systems, one often considers the use of variance reduction techniques. Iglehart and Lewis (1979) showed, using a simulation of the M/M/1 queue based on the regenerative method, that using internal linear controls reduced the variance of the estimate of the stationary waiting time of the nth customer, W_n. The linear control they identified as the most suitable reduced the standard deviation of the controlled estimate to 68% of the standard deviation of the uncontrolled estimate, equivalent to a variance reduction of .54.

In the remainder of this section, Iglehart and Lewis's results and notation are summarized and nonlinear controls for regenerative estimators are introduced in Section A.4. In Section B the expected values for several potential nonlinear controls are calculated. Section C.1 shows that for an M/M/1 queue with a traffic intensity of .5, by using a particular nonlinear control one can increase the variance reduction over a linear control . It is then shown in Section C.2 that for a traffic intensity of .99, the nonlinear controls either have bias problems or are no more effective than the linear controls.

Section D shows how one can use the regression-adjusted technique of Heidelberger and Lewis (1981) with linearly or nonlinearly controlled regenerative estimates. This combination of techniques allows one to obtain estimates of the stationary waiting time for the nth customer with much lower estimated mean square error than by using either technique alone. Examples are provided, using data from simulations of an M/M/1 queue and an M/G/1 queue, where the estimated mean square errors for the average regression-adjusted controlled estimators are 10 and 33 percent of the mean square error estimates for the crude estimators.

1. A Brief Review of the Regenerative Method

This section is condensed from Iglehart and Lewis (1979) to provide the basis for the regenerative estimator. Let $\boldsymbol{X} = \{X_t : t \geq 0\}$ be the regenerative process being simulated. Define the regenerative process as a stochastic process $\{X_t : t \geq 0\}$ where "there exists an epoch, S_1 say, such that the continuation of the process beyond S_1 is a probabilistic replica of the process beginning at time zero." (Heyman and Sobel, 1982, p. 179) Assume that $\{X_t : t \geq 0\}$ is a stable process in that $X_t \Rightarrow X$ as $t \to \infty$ where $\Rightarrow$ denotes weak convergence. One is often interested in estimating $r = \mathrm{E}[f(X)]$ for a given function $f(\cdot)$.

When using the regenerative method, one observes the pairs of random variables $\{Y_k, \tau_k : 1 \leq k \leq n\}$ where Y_k is the area under the function $f(X_t)$ in the kth cycle and τ_k is the length of the kth cycle. Two basic facts are crucial to the method. First, the successive pairs $\{Y_k, \tau_k : 1 \leq k \leq n\}$ are independent and bivariately identically distributed (i.i.d.). Second, Iglehart and Crane (1975, App. A) prove that $r = \mathrm{E}[f(X)] = \mathrm{E}[Y_1]/\mathrm{E}[\tau_1]$ using the Key Renewal Theorem (see Smith, 1958). Thus they establish that a strongly consistent point estimator for r, based on n cycles, is

$$\hat{r}(n) = \frac{\overline{Y}(n)}{\bar{\tau}(n)} \tag{86}$$

where $\overline{Y}(n) = n^{-1} \sum_{k=1}^{n} Y_k$ and $\bar{\tau}(n) = n^{-1} \sum_{k=1}^{n} \tau_k$.

There are two methods for estimating the variance of $\hat{r}(n)$. The first method involves the use of independent replications of the regenerative process, each having the same number of cycles. In this method one conducts m independent replications of the simulation, generating m independent estimates of r, namely $\hat{r}(n)_j$, for $j = 1, \ldots, m$. The sample mean and sample variance of the $\hat{r}(n)_j$ are used to calculate an overall point estimate and estimate of the variance of the point estimate. Note that one has in essence generated $N = n \times m$ total cycles and chosen to section them in a manner similar to the sectioning for quantile estimates as discussed in Section B.2 of Chapter III.

The second method of estimating the variance relies on an asymptotic expansion. Using the asymptotic expansion for the variance of a ratio, found in Cramér (1966, p. 353) or Mood, Graybill and Boes (1974, p. 181), one can write an asymptotic expansion for the variance of $\hat{r}(n)$. Letting Y represent the i.i.d. random variables Y_k and τ represent τ_k, one

can write, to order $1/n^{3/2}$,

$$\begin{aligned}
\mathrm{Var}[\hat{r}(n)] &= \mathrm{Var}\left[\overline{Y}(n)\right]\left(\frac{1}{\mathrm{E}[\tau]^2}\right) + 2\mathrm{Cov}\left[\overline{Y}(n),\bar{\tau}(n)\right]\left(\frac{-\mathrm{E}[Y]}{\mathrm{E}[\tau]^3}\right) + \mathrm{Var}[\bar{\tau}(n)]\left(\frac{\mathrm{E}[Y]^2}{\mathrm{E}[\tau]^4}\right) \\
&= \frac{1}{\mathrm{E}[\tau]^2}\left\{\mathrm{Var}\left[\overline{Y}(n)\right] - 2\left(\frac{\mathrm{E}[Y]}{\mathrm{E}[\tau]}\right)\mathrm{Cov}\left[\overline{Y}(n),\bar{\tau}(n)\right] + \left(\frac{\mathrm{E}[Y]}{\mathrm{E}[\tau]}\right)^2\mathrm{Var}[\bar{\tau}(n)]\right\} \\
&= \frac{1}{\mathrm{E}[\tau]^2}\left\{\mathrm{Var}\left[\overline{Y}(n)\right] - 2r\mathrm{Cov}\left[\overline{Y}(n),\bar{\tau}(n)\right] + r^2\mathrm{Var}[\bar{\tau}(n)]\right\}, \\
&= \frac{1}{n\mathrm{E}[\tau]^2}\left\{\mathrm{Var}[Y] - 2r\mathrm{Cov}[Y,\tau] + r^2\mathrm{Var}[\tau]\right\} \qquad (87) \\
&= \frac{1}{n\mathrm{E}[\tau]^2}\mathrm{Var}[Y - r\tau]. \qquad (88)
\end{aligned}$$

One can use (87) to estimate the variance of $\hat{r}(n)$. Substituting the sample estimates of s_Y^2, s_τ^2, $s_{Y,\tau}(n)$ and $\hat{r}(n)$ into (87) yields an asymptotic estimate of $\mathrm{Var}[\hat{r}(n)]$. Note that in the method of independent replications, the point estimate and the variance estimate are uncorrelated. However, with the asymptotic estimate of the variance (87), the estimate of the variance "is usually highly positively correlated with the point estimate." (Heidelberger and Lewis, 1981) This correlation can result in the variance estimate being artificially small when the point estimate has a small magnitude.

The asymptotic formulas (87) and (88) above do give some insight however for formulating possible control variables if one is interested in applying controls to obtain variance reduction in this regenerative simulation context. Letting $Z = Y - r\tau$, one gets that asymptotically, as the number of cycles n gets large,

$$\mathrm{Var}[\hat{r}(n)] = \frac{1}{n\mathrm{E}[\tau]^2}\mathrm{Var}[Z]. \qquad (89)$$

Equation (89) shows that the variance of $\hat{r}(n)$ is directly related to the variance of Z. Using a control variable to reduce the variance of Z will result in a reduction in the variance of $\hat{r}(n)$. As the effectiveness of a control for variance reduction is related to the correlation between the control and Z, the goal is to find a control that is strongly correlated with Z. Finding a useful control is complicated by the fact that the expected value of the control must be known either exactly or approximately. The next subsection

discusses Iglehart and Lewis's (1979) linear control for the estimate of the stationary waiting time of the nth customer in a simulated M/M/1 queue.

2. Controlling the Stationary Waiting Time of the nth Customer in an M/M/1 Queue

Define the waiting time of the nth customer in an M/M/1 queue, namely W_n, as the time from the customer's arrival until the commencement of service. One can show that under certain conditions, the waiting time process $\{W_n : n \geq 0\}$ is a regenerative process. When the queue is stable, $W_n \Rightarrow W$ as $n \to \infty$. Thus one can use a regenerative estimator similar to (86) to estimate W once suitable definitions for Y and τ are established.

In order to define Y and τ, assume the zeroth customer arrives at time $t_0 = 0$, finds the server free, and has a service time of ν_0. The nth customer arrives at time t_n and has a service time of ν_n. Define the interarrival times u_n as $u_n = t_n - t_{n-1}$ for $n \geq 1$. Assume that the ν_n and u_n sequences are independent of each other and that each consists of i.i.d. random variables. Let $\mathrm{E}[\nu_n] = \mu^{-1}$, and let $\mathrm{E}[u_n] = \lambda^{-1}$. Denote the *traffic intensity* by ρ where $\rho = \lambda/\mu$, assuming that λ and μ are both positive and μ is finite. Assume that the traffic intensity ρ is less than one so that the system is stable.

In practice, one does not have to estimate W for the M/M/1 queue as when $\rho < 1$, the expected value of W is known i.e.,

$$\mathrm{E}[W] = \frac{\rho^2}{\lambda(1-\rho)}. \tag{90}$$

However, a known value for $\mathrm{E}[W]$ provides a basis for comparing the bias of the estimate via the estimated mean square error. In later sections, the estimated bias of the controlled regenerative estimator will play a major role in assessing the effectiveness of the control scheme via the estimate of the mean square error.

Since ρ is less than one and the M/M/1 queue is stable, one can show that there exists a sequence of integer-valued random variables $\{T_k : k \geq 0\}$ such that the customers numbered T_k arrive to find the server free and experience no waiting in the queue. These customers start a new cycle or *busy period* for the system. Let $\tau_k = T_k - T_{k-1}$ for $k \geq 1$. Thus τ_k represents the number of customers served in the kth busy period (the length of

the cycle). Now define the sequence $\{Y_k : k \geq 1\}$ by

$$Y_k = \sum_{j=T_{k-1}}^{(T_k)-1} W_j, \quad \text{for } k \geq 1.$$

The random variable Y_k is sum of the waiting times in the kth busy period (the area under the function $f(\cdot)$ for the cycle).

Now that Y and τ are defined in the context of the queue, one can estimate E[W] using the regenerative estimator in (86). The estimator for E[W], based on n busy periods, is

$$\widehat{W}(n) = \frac{\overline{Y}(n)}{\bar{\tau}(n)}, \quad \text{for } n \geq 1. \tag{91}$$

One method for estimating the variance of $\widehat{W}(n)$ requires one to generate multiple, independent, replications as mentioned previously. In this case the point estimate, based on m replications of n busy periods each, would be

$$\overline{\widehat{W}}(m,n) = \frac{1}{m}\sum_{j=1}^{m} \widehat{W}_j(n). \tag{92}$$

One would use the variance of the sample mean of the $\widehat{W}_j(n)$ to estimate the variance of $\overline{\widehat{W}}(m,n)$, namely Var$\left[\overline{\widehat{W}}(m,n)\right]$ i.e.,

$$S^2_{\overline{\widehat{W}}(m,n)} = \frac{1}{m(m-1)}\sum_{j=1}^{m}\left(\widehat{W}_j(n) - \overline{\widehat{W}}(m,n)\right)^2. \tag{93}$$

The estimate of the standard deviation of $\overline{\widehat{W}}(m,n)$ would simply be the square root of $S^2_{\overline{\widehat{W}}(m,n)}$.

Another alternative for estimating the variance of the regenerative estimate of the stationary waiting time is to use the asymptotic estimate from (89). To use the asymptotic estimate of the variance for $\widehat{W}(n)$, one can use the sample estimates for the quantities in (87) i.e.,

$$s^2_{\widehat{W}(n)} = \frac{1}{n\bar{\tau}(n)}\left[s^2_Y - 2\left(\frac{\overline{Y}(n)}{\bar{\tau}(n)}\right)s_{Y,\tau}(n) + \left(\frac{\overline{Y}(n)}{\bar{\tau}(n)}\right)^2 s^2_\tau\right],$$

where s_x^2 is the sample variance of x, $s_{x,y}(n)$ is the sample covariance of x and y, and $\overline{Y}(n)/\bar{\tau}(n)$ is the sample estimate of $r = \mathrm{E}[W]$ from (91).

3. Iglehart and Lewis's Linear Control

One would like to reduce the variance of $\widehat{W}(n)$ for a given number of busy periods n. Two possibilities for reducing the variance of $\widehat{W}(n)$ are to control the Y on top of the ratio in (91) or the τ on the bottom of the ratio in (91). The controls initially considered below control the top of the ratio.

When controlling the top of the ratio in (91), one can write the controlled estimator $\widehat{W}'(n)$, as

$$\widehat{W}'(n) = \frac{\overline{Y'}(n)}{\bar{\tau}(n)} = \frac{(1/n)\sum_{i=1}^{n}\{Y_i - \theta(C_i - \mathrm{E}[C])\}}{\bar{\tau}(n)} \tag{94}$$

where C_i represents the value of an i.i.d. random variable that is the control for the ith cycle and θ is a coefficient that can be chosen so as to minimize the variance of $\widehat{W}'(n)$. One can derive an asymptotic expansion for the variance of $\widehat{W}'(n)$ as a function of θ by taking (88) and replacing $\hat{\tau}(n)$ with $\widehat{W}'(n)$ and Y by Y'. Substituting in this new expansion the value for θ that minimizes $\mathrm{Var}\left[\widehat{W}'(n)\right]$ (which can be found by differentiating the expansion with respect to θ), one can show that asymptotically

$$\mathrm{Var}\left[\widehat{W}'(n)\right] = \mathrm{Var}\left[\widehat{W}(n)\right]\left(1 - \mathrm{Cor}[C, Y - r\tau]^2\right)$$

where C represents C_k. Thus as previously implied by (89), one would like to choose a control C that is highly correlated with $Z = Y - r\tau$.

One should note several characteristics of $\widehat{W}'(n)$ that distinguish it from estimators discussed in previous chapters. The first is that the control is developed from an asymptotic formula. Thus for small sample sizes the sample correlation of the control may not translate directly into achieved variance reduction. Also, the estimator is a ratio estimator where the top and bottom elements of the ratio are highly correlated. As such it is not straightforward to predict the impact of changing or controlling either of the elements. Finally, the mean-zero control is being applied to one of the elements of the ratio, not the ratio estimator itself. Thus to assess a potential control's effectiveness, one estimates the control's correlation with the intermediate statistic Z, not the final

"statistic of interest" $\widehat{W}(n)$. The impact of these three characteristics is that unlike linear control of the mean, the R^2 from the regression that estimates the coefficients for the control function may be much different than the achieved variance reduction.

With the goal of developing controls that were highly correlated with $Z = Y - \tau\tau$, Iglehart and Lewis (1979) chose the following general form for the linear control C:

$$C = D - \tau/\mu, \tag{95}$$

where the random variable D is selected so as to "mimic" Y. Iglehart and Lewis found that dividing τ by μ in (95) helps make the variance reduction independent of the scale parameter μ.

The particular form for D in (95) favored by Iglehart and Lewis exploits a recursive relationship for calculating W_n, namely:

$$W_0 = 0, \qquad \text{and} \qquad W_{n+1} = [W_n + X_{n+1}]^+, \quad \text{for } n \geq 0,$$

where $X_n = \nu_{n-1} - u_n$ for $n \geq 1$ and the superscript plus, $^+$, denotes the plus function

$$x^+ = \begin{cases} x \text{ if } x > 0; \\ 0 \text{ otherwise.} \end{cases} \tag{96}$$

Using the X_n and the plus function, one can write Iglehart and Lewis's form for D as

$$D = \begin{cases} 0 & \tau_1 = 1, \\ X_1^+ + X_2^+ & \tau_1 \geq 2; \end{cases} \tag{97}$$

$$= \begin{cases} W_0 = 0 & \tau_1 = 1, \\ W_0 + W_1 & \tau_1 = 2, \\ W_0 + W_1 + X_2^+ & \tau_1 \geq 3. \end{cases} \tag{98}$$

The second set of equations above, (98), show that D mimics Y by using information about the waiting times of the first several customers in the busy period. In fact, if there are two or less customers for the busy period, then $D = Y$.

4. The Nonlinearly Controlled Regenerative Estimate

Given C as a potential control, one must be able to compute the expected value of C. To use C as a nonlinear control one must be able to compute the expected values for transformations of C, namely $h(C,\theta)$. Note that these transformations could be functions that act on the random variable C itself, or even functions that act on the random variables D or τ individually. As discussed in Chapter II, the transformed, nonlinear, controls $h(C,\theta)$ are incorporated into a mean-zero, linear additive function of the transformed variables and their expected values. Thus generalizing (94), one can write

$$\widehat{W}'(n) = \frac{\overline{Y'}(n)}{\bar{\tau}(n)} = \frac{\frac{1}{n}\sum_{k=1}^{n} Y_k - H(\underline{C}_k, \underline{\theta})}{\bar{\tau}(n)} \tag{99}$$

where the $H(\cdot,\cdot)$ notation comes from (10) in Chapter II.

Iglehart and Lewis (1979) derive the expected value of C, given at (95), for a queue of traffic intensity ρ, as

$$\mathrm{E}[C] = \frac{1}{\mu}\left[\frac{\rho}{1+\rho} + \left(\frac{\rho}{1+\rho}\right)^2 - \frac{1}{1-\rho}\right]. \tag{100}$$

The next several sections of the chapter will build on their results by determining the distributions and expected values for several transformations of C and demonstrating the effects of some nonlinear controls on the regenerative estimate of $\mathrm{E}[W]$.

B. CALCULATING THE EXPECTED VALUES OF POSSIBLE NONLINEAR CONTROLS

1. The Probability Function for τ

For an M/M/1 queue with traffic intensity ρ less than 1, the probability function for τ (see Kleinrock, 1975, p. 218) can be written as

$$\Pr\{\tau = l\} = \frac{1}{l}\binom{2l-2}{l-1}\rho^{l-1}(1+\rho)^{1-2l}, \quad \text{for } l \geq 1. \tag{101}$$

Note that the probability of only 1 customer in a busy period is $1/(1+\rho)$. For $\rho = .5$, the probability that there are either one or two customers in the busy period is approximately .81, while for $\rho = .99$ the probability is .63.

2. Expected Values of Transformations of D

As the distribution of τ is known, to compute the expected values for transformations of $C = D - \tau/\mu$, one must determine the distribution of C, which means that one must determine the distribution of D. One must also know the distribution of D to derive the moments of transformations of D such as D^p. One way to determine the distribution of D is to start by determining the distributions of X_1^+ and X_2^+.

The variables X_1^+ and X_2^+ both have the same type of mixed distribution. Random variables with this "lightbulb" distribution have a positive probability of being zero, but given that they are greater than zero, are distributed exponentially. By definition, X_1^+ is the remaining service time of the zeroth customer when the next customer arrives. If there is only one customer in the busy period, $\tau = 1$, then X_1^+ is zero. Thus $\Pr\{X_1^+ = 0\} = \Pr\{\tau = 1\}$ which can be seen from (101) to be $1/(1+\rho)$.

If a customer arrives before the zeroth customer finishes his service, then X_1^+ is greater than zero. This occurs with a probability of $\Pr\{\tau > 1\} = 1 - \Pr\{\tau = 1\} = \rho/(1+\rho)$. The memoryless property of the exponential service distribution implies then that given $\tau > 1$, X_1^+ is exponentially distributed with mean $1/\mu$. Thus the survivor function for X_1^+ can be written as

$$\Pr\{X_1^+ > x\} = \begin{cases} 1, & \text{for } x < 0; \\ \frac{\rho}{1+\rho} e^{-\mu x}, & \text{for } x \geq 0. \end{cases} \tag{102}$$

From (102) one can write the distribution function of X_1^+ as

$$F_{X_1^+}(x) = \Pr\{X_1^+ \leq x\} = \begin{cases} 0, & \text{for } x < 0 \\ \frac{1}{1+\rho}, & \text{for } x = 0 \\ \frac{1}{1+\rho}(1 - e^{-\mu x}), & \text{for } x > 0 \end{cases} \tag{103}$$

Since the distribution of X_1^+ has both a discrete and a continuous part, the notation $\mathrm{d}F_{X_1^+}(x)$ will be used to represent the derivative of $F_{X_1^+}(x)$ where the derivative exists, and the probability mass function for X_1^+ for those x where there is a jump in $F_{X_1^+}(x)$, namely $x = 0$.

One can determine the distribution for X_2^+ in a manner analogous to that for X_1^+. Given that X_1^+ is greater than zero, the distribution of X_2^+ is identical to that of X_1^+ since

X_2^+ is the residual service time of the $n = 1$ customer when the $n = 2$ customer arrives. If X_1^+ is zero then necessarily X_2^+ is zero. Thus one can derive the uncondition survival function for X_2^+ by multiplying the probability for $x \geq 0$ in (102) by the probability that X_1^+ is greater than 0, namely $\rho/(1+\rho)$. It follows then that

$$\Pr\left\{X_2^+ > x\right\} = \begin{cases} 1, & \text{for } x < 0; \\ \left(\frac{\rho}{1+\rho}\right)^2 e^{-\mu x}, & \text{for } x \geq 0. \end{cases} \tag{104}$$

The distribution function for X_2^+ follows directly from (104). Thus for an M/M/1 queue, X_1^+ and X_2^+ are nonnegative random variables that have different probabilities of being non-zero, but the same exponential shape parameter μ.

As the sum of two nonnegative random variables, the random variable D is also nonnegative. Thus $\Pr\{D > x\}$ is one for all x less than 0. With the survivor functions for X_1^+ and X_2^+ in (102) and (104), one can determine the survivor function for D for those x greater than zero. Let γ denote the constant $\rho/(1+\rho)$. One can write the survival function for D, for x greater than zero, using(102) and (104) as

$$\begin{aligned} \overline{F}_D(x) = \Pr\{D > x\} &= \Pr\left\{X_1^+ + X_2^+ > x\right\}, \\ &= \Pr\left\{X_1^+ > x\right\} + \int_0^x \Pr\left\{X_2^+ > (x-t) \mid X_1^+ = t\right\} \mathrm{d}F_{X_1^+}(t), \\ &= \gamma e^{-\mu x} + \gamma^2 \mu x e^{-\mu x}, \quad \text{for } x \geq 0. \end{aligned} \tag{105}$$

One must be careful when using the survivor function to remember that D has a mixed distribution with a positive probability of being equal to 0. From the definition of D in (97) and (101), it follows that $\Pr\{D = 0\} = \Pr\{\tau = 1\} = 1/(1+\rho)$. However, for x greater than 0, D has a continuous distribution. Thus for x greater than zero, one can consider the negative derivative of the survival function in (105) in terms of a "quasi" density function $\tilde{f}_D(x)$ that integrates to γ instead of one. The negative derivative of the survival function for x greater than zero can be derived from (105) as

$$\tilde{f}_D(x) = \gamma\mu e^{-\mu x} + \gamma^2\mu^2 x e^{-\mu x} - \gamma^2 \mu e^{-\mu x}, \quad \text{for } x > 0. \tag{106}$$

Now that one has the distribution function for D, one can determine the moments of transformations of D. As an example, to compute the expected value of D, namely $\mathrm{E}[D] = \int_{-\infty}^{\infty} x \mathrm{d}F_D(x)$, one can break up the expectation integral over the range from $-\infty$ to ∞. Using (106), one can compute the expected value of D as follows:

$$\begin{aligned} \mathrm{E}[D] &= \int_{-\infty}^{0} x \mathrm{d}F_D(x) + 0 \times \Pr\{D = 0\} + \int_{0}^{\infty} x \mathrm{d}F_D(x), \\ &= 0 + 0 + \int_{0}^{\infty} x \tilde{f}_D(x) \mathrm{d}x \qquad (107) \\ &= \frac{\gamma + \gamma^2}{\mu}. \qquad (108) \end{aligned}$$

This final formula for the expected value is equivalent to Iglehart and Lewis's result for $\mathrm{E}[D]$ that they determined using a conditioning argument. While Iglehart and Lewis's (1979) conditioning argument is useful for determining the expected value of D, it does not provide the survivor function for D derived in (105). One needs to know the information in the survivor function for D to compute the expected value for transformations of D.

One can compute expected values for continuous transformations of D, namely $g(D)$, in a straightforward manner. Note in (107) that the contribution to the expected value for nonpositive D is zero. Replacing the x in (107) with $g(x)$ one gets what looks like a standard integral for the expected value of a function of a continuous, positive random variable, only instead of a density function, $\tilde{f}_D(x)$ is used so that the expected value of $g(D)$ can be simplified to

$$\mathrm{E}[g(D)] = \int_{0}^{\infty} g(x) \tilde{f}_D(x) \mathrm{d}x \qquad (109)$$

One can use (109) to compute the expected value for the power transformation D^p as

$$\mathrm{E}[D^p] = \frac{1}{\mu^p} \Gamma(p+1) \left(p\gamma^2 + \gamma \right), \quad \text{for } p > -1 \qquad (110)$$

where $\Gamma(\cdot)$ represents the complete Gamma Function. For p = 1, the right side of (110) collapses to (108). The power transformation is one means of introducing nonlinearity into the control C.

Another means of introducing nonlinearity is by the creation of two new random variables through the use of a cutpoint as in Part D.2 of Chapter II. One can use D to create two new variables D_1 and D_2 through the use of a cutpoint δ as follows:

$$D_1 = \begin{cases} D & \text{if } D \leq \delta; \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad D_2 = \begin{cases} D & \text{if } D > \delta; \\ 0 & \text{otherwise.} \end{cases} \tag{111}$$

Additional nonlinearity can be introduced by applying a different power transformation to each of these new variables. One can write the expected values of D_1^p and D_2^p for $p > -1$ in terms of Incomplete Gamma functions. For D_1^p one can derive, using appropriate limits of integration in (109), that

$$\mathrm{E}[D_1^p] = \frac{\gamma}{\mu^p}\int_0^\delta \mu^{p+1}x^p e^{-\mu x}dx + \frac{\gamma^2}{\mu^p}\int_0^\delta \mu^{p+2}x^{p+1}e^{-\mu x}dx - \frac{\gamma}{\mu^p}\int_0^\delta \mu^{p+1}x^p e^{-\mu x}dx. \tag{112}$$

By exploiting the relationship between the Gamma Function and the distribution function for a Gamma (p,μ) random variable, one can simplify the notation in (112). Let $F_{\Gamma,p,\mu}(x)$ represent the cumulative distribution function of a Gamma distributed random variable with shape parameter p and mean $1/\mu$. Then (112) can be written as

$$\mathrm{E}[D_1^p] = \frac{\gamma}{\mu^p}\left[\Gamma(p+1)(1-\gamma)F_{\Gamma,p+1,\mu}(\delta) + (p+1)\gamma F_{\Gamma,p+2,\mu}(\delta)\right], \quad \text{for } p > -1. \tag{113}$$

For a given set of parameters, one can use standard software to compute the expected value in (113).

Other transformations, such as an exponential transformation, can also be used to induce nonlinearity. The expected value of $\mathrm{E}\left[e^{\eta D}\right]$ can be written as

$$\begin{aligned} \mathrm{E}\left[e^{\eta D}\right] &= \int_0^\infty \gamma e^{\eta x}\mu e^{-\mu x}dx + \gamma^2\int_0^\infty \mu^2 e^{\eta x}xe^{-\mu x}dx - \gamma^2\int_0^\infty \mu e^{\eta x}e^{-\mu x}dx, \\ &= \gamma\mu\int_0^\infty e^{-(\mu-\eta)x}dx + \gamma^2\mu^2\int_0^\infty xe^{-(\mu-\eta)x}dx - \gamma^2\mu\int_0^\infty e^{-(\mu-\eta)x}dx \\ &= \gamma\left(\frac{\mu}{\mu-\eta}\right)^2 + \gamma^2\left(\frac{\mu}{\mu-\eta}\right)^2 - \gamma^2\left(\frac{\mu}{\mu-\eta}\right). \end{aligned} \tag{114}$$

As the distribution of D is fairly tractable, other transformations of D could be used if desired *as long as the expected value of the transformed D can be computed.*

To determine the expected value for transformations of $C = D - \tau/\mu$, now that the distribution for D is known, one could work directly with the distributions for D and τ. However this gets messy as D and τ are not independent. It is easier to work with the conditional distribution for C given τ and then uncondition. Sections B.3, B.4 and B.5 will cover determining the distribution of C. Section B.6 will develop formulas for the expected value of transformations of C.

3. The Conditional Survivor Function for D

One manner of determining the expected value of transformations of C involves working with the conditional survivor function for D given τ. As shown just prior to (105), the probability that $D > x$, for $x < 0$, is 1 for all τ. For $x > 0$, one can derive, using (105), that

$$\begin{aligned} \Pr\{D > x \mid \tau = 1\} &= 0, \\ \Pr\{D > x \mid \tau = 2\} &= e^{-\mu x}, \text{ and} \\ \Pr\{D > x \mid \tau = 3\} &= e^{-\mu x} + \int_0^x e^{-\mu(x-t)} \Pr\left\{X_2^+ > 0 \mid \tau = 3\right\} \mu e^{-\mu t} dt, \\ &= e^{-\mu x} + (1/2)\mu x e^{-\mu x}, \\ &= e^{-\mu x}[1 + (1/2)\mu x]. \end{aligned} \tag{115}$$

Following in a similar manner, one can derive a more general expression, for $\tau = l \geq 2$, as

$$\Pr\{D > x \mid \tau = l \geq 2\} = \begin{cases} 1 & \text{for } x \leq 0; \\ e^{-\mu x}[1 + c_l \mu x], & \text{for } x > 0 \end{cases} \tag{116}$$

where c_l is the probability that X_2^+ is greater than 0 given $\tau = l$. It is shown next that $c_l = (3l - 6)/(4l - 6)$.

4. Determining c_l

a. *Determining* $\Pr\{X_1^+ > 0 | \tau\}$ *and* $\Pr\{X_2^+ > 0 | \tau\}$

For both X_1^+ and X_2^+, the probability of being greater than zero can be considered as a function of τ, the number of customers in the busy period. This allows one to work with the conditional distributions for X_1^+ and X_2^+ given $\tau = l$. As discussed

earlier, if the busy period has only one customer, $\tau = 1$, this implies that X_1^+ and X_2^+ are identically zero since

$$(\tau = 1) \text{ implies } (\nu_0 < \mu_1), \text{ which implies } (X_1^+ = 0), \text{ which implies } (X_2^+ = 0).$$

Therefore,

$$\Pr\left\{X_1^+ > 0 \mid \tau = 1\right\} = 0 \quad \text{and} \quad \Pr\left\{X_2^+ > 0 \mid \tau = 1\right\} = 0.$$

If the busy period has at least two customers, $\tau = l \geq 2$, then

$$\Pr\left\{X_1^+ > 0 \mid \tau = l \geq 2\right\} = 1$$

since for there to be two or more customers in a busy period,

$$(\nu_0 > \mu_1), \text{ which implies } (X_1 > 0) \text{ which implies } (X_1^+ > 0).$$

However, having two or more customers in a busy period does *not* imply that X_2^+ is positive; it may be zero. In what follows it will be shown that for $\tau \geq 2$,

$$c_l = \Pr\left\{X_2^+ > 0 \mid \tau = l \geq 2\right\} = \frac{3l - 6}{4l - 6}.$$

b. *Determining* $\Pr\{X_2^+ > 0 \mid \tau = l \geq 2\}$

This subsection considers a single busy period that starts at time 0 and has exactly $\tau = l$ customers for some $l \geq 1$. Let $L(t)$ represent the number of customers in the system at time t with $L(0) = 0$. Call the time at which a customer arrives or departs a *transition time* for $L(t)$ since $L(t)$'s value changes by plus or minus one respectively. Denote the jth transition time in a busy period as t_j where j ranges from 0 to 2τ.

One can consider the process $L(t_j)$ as a discrete-time Markov Chain that for a given τ may have several different sample paths, only some of which have $X_2^+ > 0$. For example, if there are exactly three customers in the busy period, $\tau = 3$, there are two possible sample paths for the number of customers in the system as t_j goes from 0 to 6: $L(t_j) = 0,1,2,3,2,1,0$ and $L(t_j) = 0,1,2,1,2,1,0$. Only the first of these paths has X_2^+ greater than zero.

The conditional probability that X_2^+ is greater than 0 given τ is simply the total number of paths for which X_2^+ is greater than 0 divided by the total number of paths given τ, where this total includes both the paths for which X_2^+ is greater than 0 and the paths for which X_2^+ equals zero. To calculate $\Pr\{X_2^+ > 0 \mid \tau\}$, one must first determine the total number of paths and then determine the total number of paths for which $X_2^+ > 0$.

Determining the total number of paths for a given value of τ is straightforward. Examining the probability function for the number of customers in a busy period of M/M/1 queue with traffic intensity ρ in (101), one can see that the $(1/l)\binom{2l-2}{l-1}$ terms provide the total number of sample paths given $\tau = l$, and the ρ terms are the probability of a path of length $2l$. The terms can be broken out this way because of the Markovian nature of the transitions in an M/M/1 queue as will now be explained.

Since each sample path is independent, the probability that $\tau = l$ can be calculated as the probability of having a sample path of length $2l$ times the number of sample paths of length $2l$. Consider a busy period that has exactly l customers. It must then have a sample path of length $2l$. To start the busy period there must be an arrival. In order for there to be more than one customer in the busy period, the next arrival must occur prior to Customer 1's departure. Because of the Markovian property of the M/M/1 queue, this occurs with probability $\lambda(\mu + \lambda) = \rho(1 + \rho)^{-1}$. In order for there to be at least l customers in the busy period there must be at least $l - 1$ more instances where the next transition is an arrival rather than a departure. In order for there to be no more than l customers, there must be l instances where the next transition is a departure rather than an arrival. This occurs with probability $\mu/(\mu + \lambda) = (1 + \rho)^{-1}$, again because of the Markovian property of the queue. The instances where the next arrival is a departure rather than arrival may be interspersed between the arrivals as long as the number of arrivals is greater than the number of departures or it reaches l. Thus the length of the sample path is $1 + (l - 1) + 2l$, 1 for the initial arrival, the $l - 1$ for the arrivals prior to departures and the l for the departures prior to arrivals. The probability of having the $l - 1$ arrivals is $\rho^{l-1}(1 + \rho)^{1-l}$. The probability of having the l departures is $(1 + \rho)^{-l}$. Thus, since the arrivals and departure times are independent, the probability of having a sample path of length $2l$ is $\rho^{l-1}(1+\rho)^{1-2l}$. The remaining $(1/l)\binom{2l-2}{l-1}$ terms in (101) account for the number of paths of length $2l$.

Let $NP(l)$ denote the total number of sample paths given $\tau = l$. It follows then that

$$NP(l) = \frac{1}{l}\binom{2l-2}{l-1}, \quad \text{for } l \geq 1. \tag{117}$$

It will be useful later on to have an expression for $NP(l)$ in terms of $NP(l-1)$. Using (117), one can easily derive the recursive expression

$$NP(l) = \left[\frac{2(2l-3)}{l}\right] NP(l-1), \quad \text{for } l \geq 2. \tag{118}$$

Now that the total number of paths can be calculated, one must determine the total number of paths for which X_2^+ is greater than zero. Drawing a graph of the possible sample paths helps to determine the number of paths for which X_2^+ is greater than zero. By plotting possible sample paths of $L(t_j)$ against t_j for a given τ, one can create a directed graph that contains all of the possible sample paths from $L(0) = 0$ to $L(t_{2\tau}) = 0$.

Figure 32 shows the possible sample paths for $\tau = 3$. The y (vertical) axis is $L(t_j)$, the number of customers in the system at time t_j. The x (horizontal) axis is the transition number i.e., the index j of t_j for $j = 0$ to 2τ. A point is represented by the ordered pair (a,b) where a is the transition number from the x axis and b is the value of $L(t_j)$.

The lines connecting the points represent the transitions; a $+1$ slope is an arrival and a -1 slope is a departure. Slopes other than ± 1 are not allowed since there are only the two types of transitions. The arrows on the lines indicate the directed nature of the graph in that as the x axis represents subsequent points in time, one can only move to the right.

A sample path from a point (a,b) to another point (c,d) is a set of points such that the first element of the ordered pairs is strictly increasing from a to c and the second element of each pair is either 1 greater or 1 less than the second element in the preceding pair. A sample path from $(3,1)$ to $(6,0)$ in Figure 32 is the set $\{(3,1),(4,2),(5,1),(6,0)\}$.

In short, there are several constraints in constructing a graph of the possible sample paths for the number of customers in the system for a busy period with τ customers.

1. The graph always starts at the point $(0,0)$.

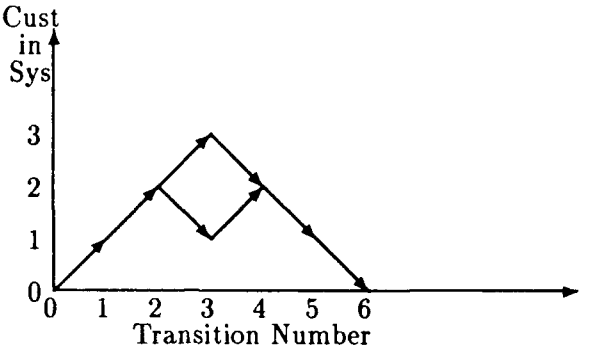


Figure 32. Graph of the possible sample paths for a busy period with 3 customers.

2. The value of $L(t_j)$ is less than or equal to t_j for $t_j \leq \tau$, and is less than or equal to $2\tau - t_j$ for $t_j > \tau$.
3. The point $(a,0)$ only exists for $a = 0$ and $a = 2\tau$.
4. The lines must have slopes of ± 1.

To construct a graph for a given τ, one starts at the point $(0,0)$ and creates all transitions permitted by the above constraints. When $\tau = 3$, the constraints lead to the creation of the graph in Figure 32. The two possible sample paths from $(0,0)$ to $(2\tau,0)$ for $\tau = 3$ in Figure 32 are

$$\{(0,0),(1,1),(2,2),(3,3),(4,2),(5,1),(6,0)\}$$

and

$$\{(0,0),(1,1),(2,2),(3,1),(4,2),(5,1),(6,0)\}.$$

The line connecting $(2,2)$ to $(3,3)$ in Figure 32 must be part of the path for X_2^+ to be greater than 0. To determine the number of paths for which $X_2^+ > 0$, one can count the paths from $(3,3)$ to $(2\tau = 6,0)$. Counting is easy in this graph since there is only one path. Note that one can determine the number of paths for which $X_2^+ = 0$ (again 1 path) by counting the paths from $(3,1)$ to $(2\tau,0)$.

As τ increases, the number of paths increases and the graph gets larger. Figure 33 shows the five possible sample paths from $(0,0)$ to $(8,0)$ for a busy period with

4 customers. Again one can count the paths for which $X_2^+ > 0$ (3 of them) by counting the paths from (3,3) to ($2\tau = 8, 0$) and one can count the number of paths for which $X_2^+ = 0$ (2 of them) by counting the paths from (3,1) to (8,0). Counting paths gets tedious as τ increases since, as can be seen from (118), as l increases by 1, the number of paths goes up by a factor that approaches 4 as $\tau = l$ gets large; for $\tau = 8$, the total number of paths is 429.

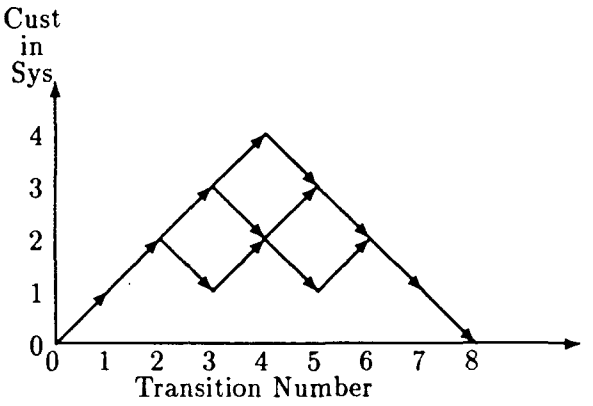


Figure 33. Graph of the possible sample paths for a busy period with 4 customers.

Comparing the graphs for $\tau = 3$, Figure 32, and for $\tau = 4$, Figure 33, one can see that the number of paths from (3,1) to (8,0) in Figure 33 is identical to the number of paths from (1,1) to (6,0) in Figure 32. Thus for $\tau = 4$, the number of paths for which $X_2^+ = 0$ equals the total number of paths for $\tau = 3$. This can be shown to be true in general. Let $NP_{X_2^+=0}(l)$ denote the total number of paths for which $X_2^+ = 0$ for a given $\tau = l$.

Lemma 3 *For $\tau = l$ where $l \geq 3$, the total number of paths such that $X_2^+ = 0$ equals the total number of paths for $\tau = l - 1$.*

Proof

The proof that follows depends upon constructing the graph for $\tau = l$ from the graph for $\tau = l - 1$ in a special way. There are two methods for expanding a graph for $\tau = l - 1$ to a graph for $\tau = l$. The first method expands to the right by adding the set of l

points $\{(l,l),(l+1,l-1),(l+2,l-2),\ldots,(2l,0)\}$, then adding the permissible connecting lines and finally deleting the line from $(2(l-1)-1,1)$ to $(2(l-1),0)$.

The second, and more useful method for the purpose at hand, is to expand the graph to the left and then shift the x axis.

1. First add the set of points $\{(l-2,l),(l-3,l-1),(l-4,l-2),\ldots,(-2,0)\}$ and delete the line from $(0,0)$ to $(1,1)$. Figure 34 shows the results of these operations on a graph where $l-1=3$. As a result of expanding to the left, no paths are created or destroyed from $(1,1)$ to $(2(l-1),0)$; thus the total number of paths from $(1,1)$ to $(2(l-1),0)$ remains $NP(l-1)$.
2. Now shift the x axis to the right by adding 2 to each transition number. The former point $(1,1)$ becomes the point $(3,1)$ and the former point $(2(l-1),0)$ becomes $(2l,0)$ and again no paths have been created or destroyed between these two points.
3. Now the total number of paths for which $X_2^+ = 0$ in the graph for $\tau = l$ is the same as the total number of paths in the original graph for $\tau = l-1$. By adding in the permissible lines for the new points on the left, the graph contains all the sample paths for a busy period with $\tau = l$ customers.

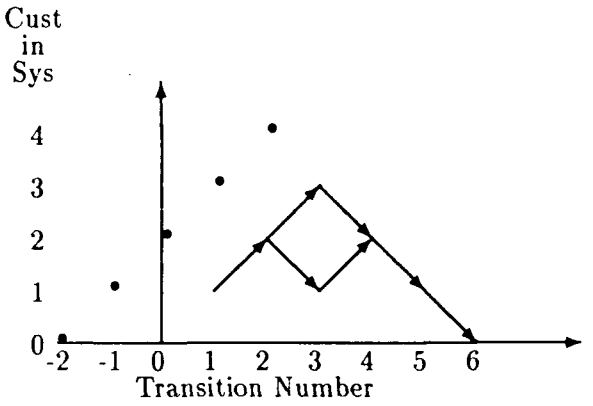


Figure 34. The results of expanding a graph for $l-1=3$ to the left.

Given that $NP_{X_2^+=0}(l) = NP(l-1)$, it follows that

$$\Pr\left\{X_2^+ > 0 \mid \tau = l \geq 3\right\} = \frac{NP(l) - NP_{X_2^+=0}(l)}{NP(l)},$$

$$= 1 - \frac{NP(l-1)}{NP(l)},$$

$$= 1 - \frac{l}{2(2l-3)},$$

$$= \frac{3l-6}{4l-6}, \quad \text{for } l \geq 3.$$

Since $\Pr\left\{X_2^+ > 0 \mid \tau = 2\right\} = 0$, it is also true for $l \geq 2$ that

$$c_l = \Pr\left\{X_2^+ > 0 \mid \tau = l\right\} = \frac{3l-6}{4l-6} \tag{119}$$

as was to be shown.

By manipulating (119), it is easy to see that as l gets large c_l approaches 3/4. Thus regardless of how large τ may be, X_2^+ does not have to be positive i.e., it can be zero.

5. The Distribution of $C = D - \tau/\mu$

Now that c_l can be computed, the conditional survivor function for D for $\tau \geq 2$ in (116) is completely determined. Given (115), (116), and the formula for c_l in (119), one can derive the survivor function for C by conditioning on τ as follows:

$$\begin{aligned}
\Pr\{C > x\} &= \Pr\{(D - \tau/\mu > x)\}, \\
&= \sum_{l=1}^{\infty} \Pr\{D > (x + \tau/\mu) \mid \tau = l\} \Pr\{\tau = l\}, \\
&= \sum_{l=1}^{\infty} \Pr\{D > (x + l/\mu) \mid \tau = l\} \Pr\{\tau = l\}.
\end{aligned}$$

For $\tau = l = 1$, using (115),

$$\Pr\{D > (x + 1/\mu) \mid \tau = 1\} = \begin{cases} 1, & \text{for } (x + 1/\mu) < 0; \\ 0, & \text{for } (x + 1/\mu) \geq 0 \end{cases} \tag{120}$$

since $\tau = 1$ implies that D is identically zero. For $\tau = l \geq 2$, it follows from (116) that

$$\Pr\{D > (x + l/\mu)\} = \begin{cases} 1, & \text{for } (x + l/\mu) < 0; \\ e^{-\mu(x+l/\mu)} \left[1 + c_l \mu \left(x + \frac{l}{\mu}\right)^+\right], & \text{for } (x + 1/\mu) \geq 0. \end{cases} \tag{121}$$

By unconditioning on τ and combining the above expressions, (120) and (121), one can write the unconditional survivor function for C, for $-\infty < x < \infty$, as

$$\Pr\{C > x\} = \mathrm{I}\left[(x+1/\mu)<0\right]\frac{1}{(1+\rho)} + \sum_{l=2}^{\infty} e^{-\mu(x+l/\mu)^+}\left[1 + c_l\mu(x+l/\mu)^+\right]\Pr\{\tau = l\} \quad (122)$$

where $\Pr\{\tau = l\}$ comes from (101) and $I[T]$ is an indicator function that is 1 if T is a true statement and 0 otherwise.

Looking closely at (122), one can see that the distribution for C is a mixed distribution with a positive probability for $C = -1/\mu$. The need for plus functions to describe the mixed distribution of C makes (122) not all that useful for determining the expected values of transformations of C. It is much easier to first condition on τ, then determine the expected value of a transformed C, and then uncondition.

6. Formulas for the Expected Value of Transformations of C

Determining the expected value of C and transformations of C using (122) can be difficult because of the plus functions. By conditioning on τ, one can take advantage of the fact that $\mathrm{E}[f(C)] = \mathrm{E}[\mathrm{E}[f(C) \mid \tau]]$ and work with the conditional distribution for C. The first step in determining expected values is to determine $\mathrm{d}F_{C|\tau}(x)$.

Working with (115), (116), and (122), one can derive the following expressions:

$$\mathrm{d}F_{C|\tau=1}(x) = \mathrm{I}\left[x = -1/\mu\right], \quad (123)$$

$$\mathrm{d}F_{C|\tau=2}(x) = \begin{cases} 0, & \text{for } x \le -2/\mu; \\ \mu e^{-\mu(x+2/\mu)}, & \text{for } x > -2/\mu \end{cases} \quad (124)$$

and

$$\mathrm{d}F_{C|\tau=l\ge 2}(x) = \begin{cases} 0, & \text{for } x \le -l/\mu; \\ \mu(1-c_l)\mu e^{-\mu(x+l/\mu)} + c_l\mu^2(x+l/\mu)e^{-\mu(x+l/\mu)}, & \text{for } x > -l/\mu. \end{cases} \quad (125)$$

Note that (124), which is actually a special case of (125), is displayed for clarity.

One can determine the conditional expected values for C given τ by integrating (123) and (125) over the ranges where $\mathrm{d}F_{C|\tau}(x)$ is non-zero. Doing these operations, with the change in variable of integration of $y = x + l/\mu$, yields the following conditional

expected values:

$$\mathrm{E}[C \mid \tau = 1] = \frac{-1}{\mu}, \tag{126}$$

$$\mathrm{E}[C \mid \tau = 2] = \frac{-1}{\mu}, \text{ and} \tag{127}$$

$$\mathrm{E}[C \mid \tau = l \geq 2] = \frac{-1}{\mu}(l - 1 - c_l) \tag{128}$$

Unconditioning then entails summing the conditional expected values from (126) and (128) over the possible values for τ while weighting them by the probabilities for $\tau = l$ i.e.,

$$\begin{aligned}
\mathrm{E}[C] &= \sum_{l=1}^{\infty} \mathrm{E}[C \mid \tau = l] \Pr\{\tau = l\}, \\
&= \frac{-1}{\mu} \Pr\{\tau = 1\} + \sum_{l=2}^{\infty} \frac{-1}{\mu}(l - 1 - c_l) \Pr\{\tau = l\}, \\
&= \left(\frac{-1}{\mu}\right)\left(\frac{1}{1+\rho}\right) + \sum_{l=2}^{\infty} \frac{-1}{\mu}(l - 1 - c_l)\frac{1}{l}\binom{2l-2}{l-1}\rho^{l-1}(1+\rho)^{1-2l}
\end{aligned} \tag{129}$$

This is the final result. While no closed-form solution is apparent for (129), it is possible to determine $\mathrm{E}[C]$ directly from $\mathrm{E}[D]$ and $\mathrm{E}[\tau]$ as Iglehart and Lewis (1979) did in calculating (100). Their method though provides only the expected value of C. The expression for $\mathrm{E}[C]$ in (129) is useful in that it provides the basis for determining the expected values of transformations of C.

One can use (123) and (125) to derive an expression for the expected value of transformations of C i.e., $g(C)$ where $g(x)$ is a continuous transformation that is well defined for $-\infty < x < \infty$. To demonstrate the method for determining an expression for $\mathrm{E}[g(C)]$, let $g(x)$ be x^p. Since C can be negative or zero, the power parameter p must be restricted to being a positive even integer. Using (123) and (125), one can write the general expressions

$$\mathrm{E}[C^p \mid \tau = 1] = \left(\frac{-1}{\mu}\right)^p \tag{130}$$

and

$$\begin{aligned} \mathrm{E}[C^p \mid \tau = l \geq 2] &= \int_{-l/\mu}^{\infty} \left\{ (1-c_l) x^p e^{-\mu(x+l/\mu)} + c_l \mu^2 x^p (x+l/\mu) e^{-\mu(x+l/\mu)} \right\} \mathrm{d}x, \\ &= \int_{0}^{\infty} \left\{ (1-c_l)(y-l/\mu)^p e^{-\mu y} + c_l \mu^2 (y-l/\mu)^p y e^{-\mu y} \right\} \mathrm{d}y. \end{aligned} \quad (131)$$

After integrating and a bit of manipulation, one can write the conditional expected value in (131) as

$$\mathrm{E}[C^p \mid \tau = l \geq 2] = \left(\frac{1}{\mu}\right)^p \left\{ (1-c_l) \sum_{j=0}^{j=p} (-1)^j \binom{p}{j} l^{p-j} j! + c_l \sum_{j=0}^{j=p} (-1)^j \binom{p}{j} l^{p-j} (j+1)! \right\} \quad (132)$$

where p is an even integer and c_l is from (119).

To get the unconditional expected value, one must now sum (131) and (132) with respect to the probability mass function for τ. Thus one can write the expected value of C^p for positive even-integer p as

$$\begin{aligned} \mathrm{E}[C^p] &= \left(\frac{-1}{\mu}\right)^p (\Pr\{\tau = 1\}) \\ &+ \left(\frac{1}{\mu}\right)^p \sum_{l=2}^{l=\infty} \sum_{j=0}^{j=p} \left\{ (1-c_l)(-1)^j \binom{p}{j} l^{p-j} j! + c_l (-1)^j \binom{p}{j} l^{p-j} (j+1)! \right\} \Pr\{\tau = l\} \end{aligned} \quad (133)$$

Depending upon the probability mass function for τ, no closed-form solution may be apparent. In some cases, one can use direct computation to get an approximate answer.

Now that expected values of some nonlinear controls can be obtained, the next section will use a simulation to evaluate the performance of these controls in reducing the variance and mean square error of the regenerative estimate of the stationary waiting time an M/M/1 queue.

C. NONLINEAR CONTROLS FOR THE STATIONARY WAITING TIME IN AN M/M/1 QUEUE

A simulation experiment was conducted to determine the performance of several nonlinear controls at reducing the variance and mean square error of the regenerative estimate of the stationary waiting time in the M/M/1 queue. The major factors in the simulation experiment were the traffic intensity ρ, the form of the nonlinear control and the number of cycles, n, used to compute $\widehat{W}'(n)$ using (94).

The initial traffic intensity was $\rho = .5$. It was then increased to $\rho = .99$. The standard for judging effectiveness was the linear control of $C = d - \tau/\mu$ from (95) where D comes from (97). Several nonlinear controls were examined for $\rho = .5$ and only the most effective were examined at $\rho = .99$.

The estimators used in the experiment are as follows:

- $\overline{\widehat{W'}}(m,n)$ is the controlled estimate of the stationary waiting time. The m and n indicate that m independent estimates $\widehat{W}'_j(n)$, for $j = 1,\ldots,m$, from (94), were averaged to get the estimate as in (92).
- $s^2_{\overline{\widehat{W'}}(m,n)}$ is the estimate of the variance of $\overline{\widehat{W'}}(m,n)$. This is calculated as variance of the m estimates $\widehat{W}'_j(n)$ divided by m as in (93). The standard deviation (SD) is the square root of $s^2_{\overline{\widehat{W'}}(m,n)}$.
- The mean square error (MSE) is estimated as the sum of $s^2_{\overline{\widehat{W'}}(m,n)}$ and $(\overline{\widehat{W'}}(m,n) - \mathrm{E}[W])^2$ where $\mathrm{E}[W]$ is the known expected value for W given in (90).

The experiment showed that a nonlinear control could be more effective than the linear control at a traffic intensity of .5. However, when the traffic intensity was increased to .99, the nonlinear controls were either no more effective than the linear control or had too much bias to be useful.

1. The M/M/1 Queue with Traffic Intensity of .5

The first part of the experiment consisted of simulating an M/M/1 queue with a traffic intensity of .5. The interrarrival and service rates were chosen so that the traffic intensity was .5 and the expected value for W, computed using (90), was 10. The M/M/1 queue was simulated until 120,000 total busy periods or cycles were completed. The data was collected for calculating Y, τ and D for each busy period.

Breiman and Friedman's (1985) ACE program, discussed in Chapter II, was used to initially assess the performance of the linear control vis-a-vis the best nonlinear control, the control being a function of C. Iglehart and Lewis (1979) reported that the linear control reduced the standard deviation of the crude estimate by about 70%, or equivalently, achieved an R^2 of .54. When the ACE program was given $C = D - \tau/\mu$ as the independent variable and $Z = Y - r\tau$ as the dependent variable, the estimate of R^2 was between .58 and .76 depending upon the sample. The fact that the estimated R^2s from ACE were higher

than that obtained by the linear control indicated that using a nonlinear control instead of the linear control could increase the obtained variance reduction.

Figure 35 is a bivariate scatterplot, using data from one sample of 5,000 busy periods, which typifies the untransformed relationship between $Z = Y - r\tau$, on the y axis, and the linear control $C = D - \tau/\mu$ on the x axis. When looking at the graph one sees less than 5,000 points. This is due to two reasons. The first is the printer resolution. The second is the fact that a single dot represents all the busy periods that have only one customer; for those busy periods the control value for the x axis is $0 - 1/\mu$ and the Z value for the y axis is $0 - r$. This single dot can represent a large number of points since the probability that a busy period has just one customer is $1/1 + \rho$. For $\rho = .5$ the probability that a busy period has only one customer is .66. One can see in Figure 35 that there is a "cloud" of points in the lower right representing the busy periods with few customers. The highly scattered points on the left side of the graph represent the busy periods where there are many customers.

Figure 36 shows the transformation ACE determined for maximizing the correlation between the Z data and the transformed control data in Figure 35. In this figure the y axis contains the transformed C values and the x axis contains the original C values. On the left side of Figure 36 the transformation looks linear. However, on the right side of the graph, where the transformation curves back up, the approximating transformation is clearly nonlinear.

In general,the graphs of the ACE approximating transformations for C had the distinct parabolic style curvature shown in Figure 36. In an effort to get a transformation of $C = D - \tau/\mu$ to mimic the parabolic shape there were essentially two approaches. The first approach was to transform C. The second approach was to transform D individually. Since D is a nonnegative random variable and its distribution is more tractable than that of C, more transformations of D were possible. In short, the following nonlinear controls were compared to the baseline linear control, $C = D - \tau/\mu$, for controlling the top of the ratio estimator (94):

1. $h(C) = C + C^2$,
2. $h(C, p) = \frac{D^p - 1}{p} - \tau/\mu$,

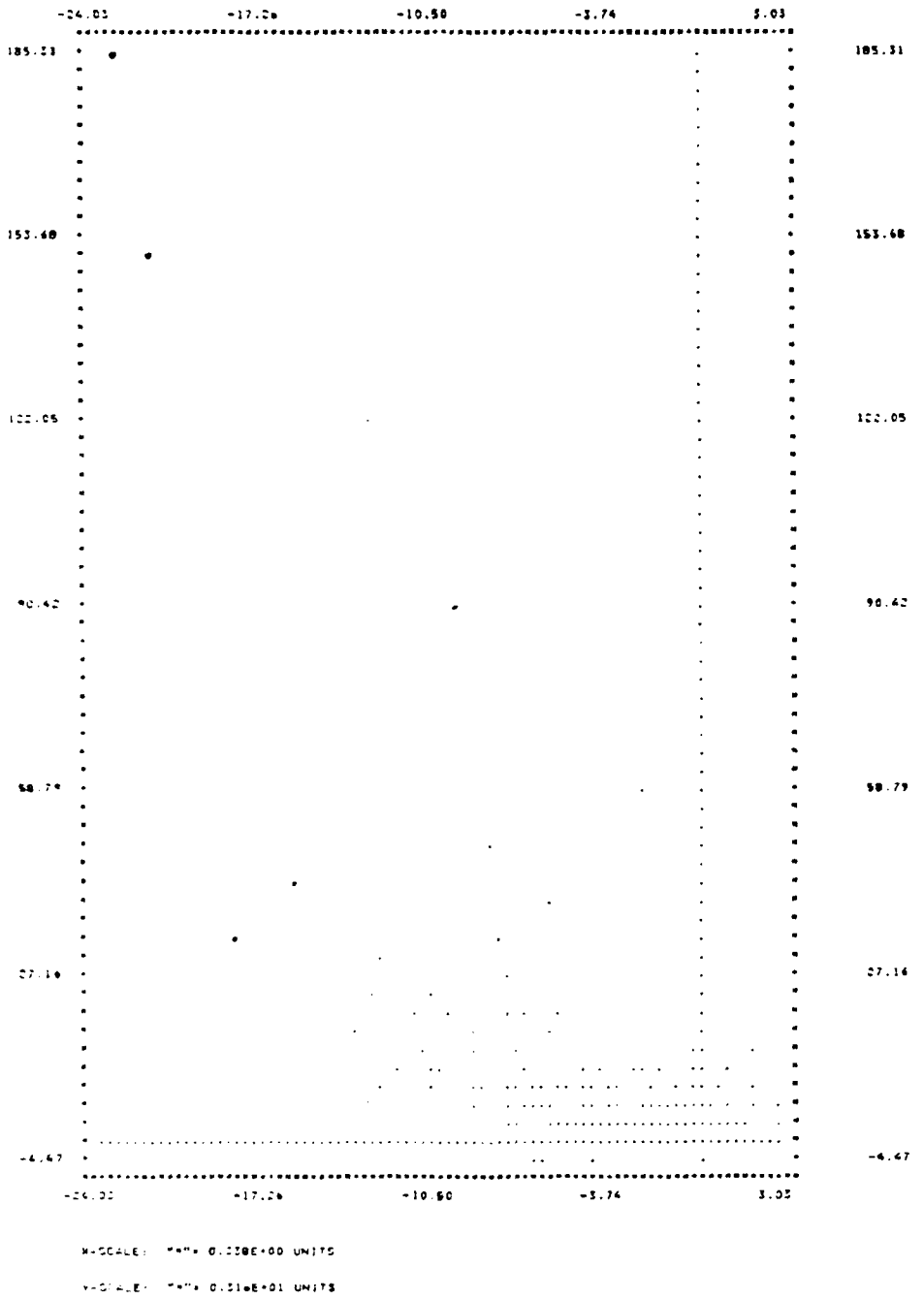


Figure 35. A bivariate scatterplot of Z versus $C = D - \tau/\mu$ for a sample of 5,000 busy periods of an M/M/1 queue with ρ=.5.

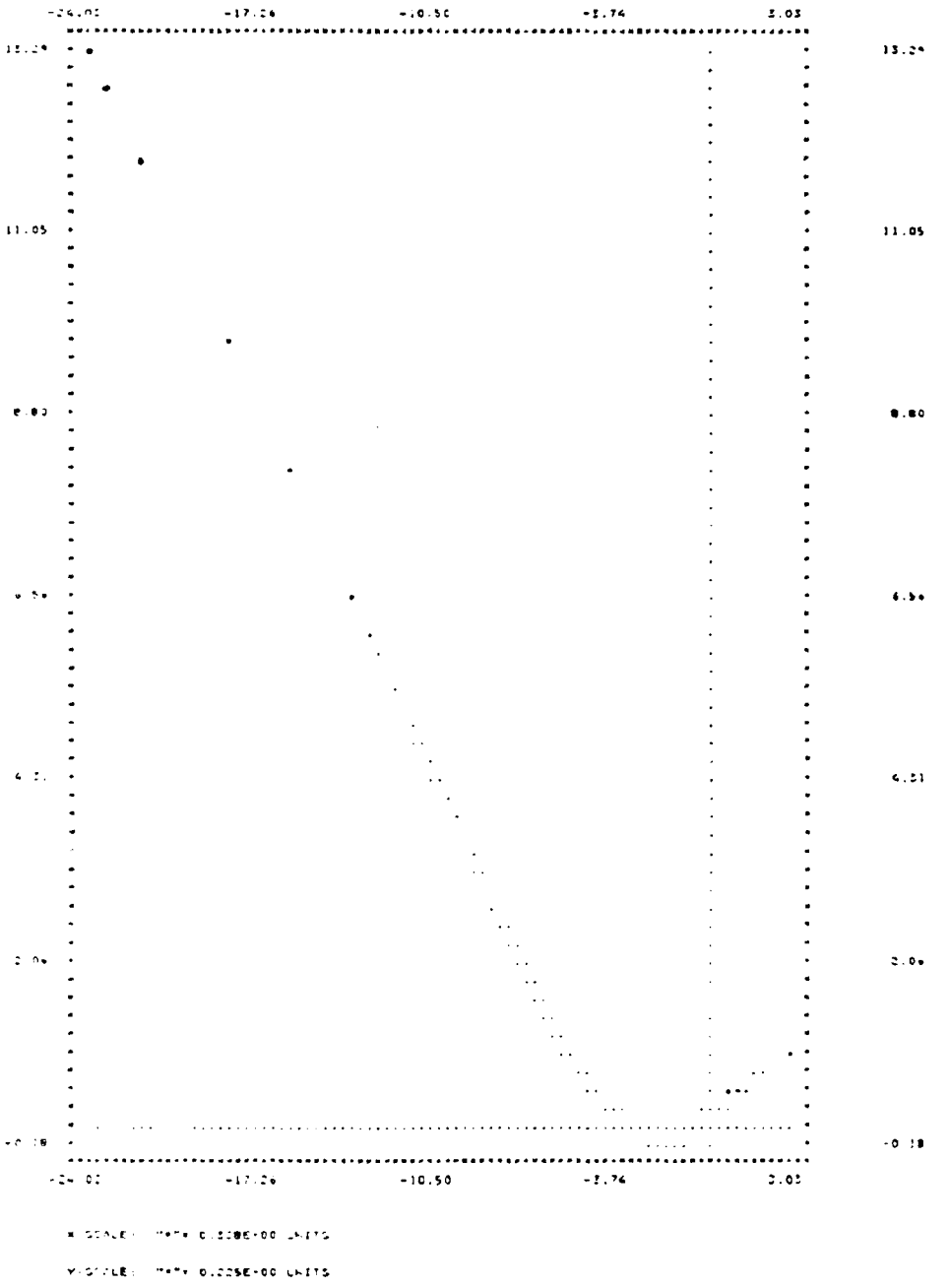


Figure 36. A bivariate scatterplot of the ACE transformed C versus $C = D - \tau/\mu$ for a sample of 5,000 busy periods of an M/M/1 queue with ρ=.5.

3. $h(\underline{C},p) = (D - \tau/\mu) + \left(\frac{D^p-1}{p} - \tau/\mu\right)$, and

4. $h(C,\underline{p}) = \left(\frac{D_1^{p_1}-1}{p} - \tau/\mu\right) + \left(\frac{D_2^{p_2}-1}{p} - \tau/\mu\right)$

where D_1 and D_2 in the fourth control are formed as in (111). Control 1 is actually still a linear control (although a multiple linear control) as the exponent was fixed at 2 beforehand. Control 2 is the simplest of the nonlinear controls, with Controls 3 and 4 becoming more complex.

The four nonlinear controls above were individually used in the control function as in (99) to calculate a nonlinearly controlled regenerative estimate. Table 3 contains the estimates of the standard deviation associated with the crude estimate, the linearly controlled estimate and the four nonlinearly controlled estimates. Each estimate was based on the same sample of 120,000 busy periods or cycles. The table shows the standard deviation estimate by type estimator and by the number of busy periods, n, used to compute $\widehat{W}_j'(n)$. The number in () is the ratio of the controlled standard deviation to the crude standard deviation. This is the measure of variance reduction used by Iglehart and Lewis (1979).

Table 3 shows several things. First, for this estimate, the linear control did better than expected. For n=4000, the standard deviation of the linear control estimate was 57% of the crude estimate as compared to the 68% reported by Iglehart and Lewis (1979). Among the nonlinear controls, the second control above had the smallest estimates of the standard deviation; for n=2000 $s_{\widehat{W}'(m,n)} = .059$ and for $n = 4000$ $s_{\widehat{W}'(m,n)} = .05816$. These estimates translate to variance reduction R^2's of .73 and .76 respectively, values which are close to the maximum ACE estimates.

One can also see in Table 3 that the addition of C^2 to the linear control function did *not* improve the estimates over the linear control C by itself; in fact the estimated standard deviation was larger for n = 500, 1,000 and 2,000. This lack of improvement may be due in part to the variance inflation that results from having to estimate a second coefficient as discussed in Section B.3 of Chapter II. As will be discussed again in Section A.2.a, a more important reason is probably the impact of the small sample bias of the control function increasing the bias of the controlled estimate. The tendency of a control function to affect the bias of the controlled estimate is a function of the variance of the control function. Since Control 1 contains C^2 in addition to C, it has much higher variance

than the baseline linear control of just C. Thus Control 1 is much more likely to be further from its asymptotic mean of zero than the baseline linear control. This small sample bias can affect the bias of the controlled estimate. The point is that one would predict that Control 1, with its additional term, would be more effective than the baseline linear control and in fact Control 1 is less effective. This is one of the characteristics of the controlled ratio estimator discussed in Section A.3.

In summary, Table 3 shows that based on a single estimate, the more complicated Controls 3 and 4 did no better than Control 2, and Control 2 reduced the standard deviation more than the linear control.

Control		Standard Deviation (% of Crude)			
	Estimator	n=500	n=1000	n = 2000	n=4000
	Crude	.11698	.11366	.11390	.11987
	Lin Con	.07117 (60)	.07489 (66)	.06885 (60)	.06815 (57)
1	$C + C^2$	.08157 (72)	.08585 (76)	.0794 (68)	.06628 (55)
2	D^p	.06214 (52)	.06543 (57)	.05900 (52)	.05816 (49)
3	$D + D^p$	.06340 (54)	.06736 (59)	.06013 (52)	.06355 (53)
4	$D_1^{p_1}$	.06186 (52)	.06236 (55)	.06011 (52)	.06748(56)

TABLE 3. Estimates of the standard deviation of $\widehat{\overline{W}}'(m,n)$ for various n from 120,000 busy periods of an M/M/1 queue with traffic intensity of .5. The number in () is 100 times the ratio of the estimated controlled standard deviation to the crude estimated standard deviation.

In order to evaluate the bias characteristics of the estimates as well as their standard deviations, Table 4 shows the estimated mean square errors associated with the standard deviation estimates in Table 3. Again Control 2, D^p does better than the linear control at reducing the mean square error and none of the more complex controls do any better. Control 1, $C + C^2$, actually increases the mean square error due to severe bias problems in $\widehat{\overline{W}}'(m,n)$. As n increases to 4,000, the inflation lessens but does not go away.

Tables 3 and 4 show an estimate of the standard deviation and mean square error for each control and n combination. To get estimates of the precision of the estimates of

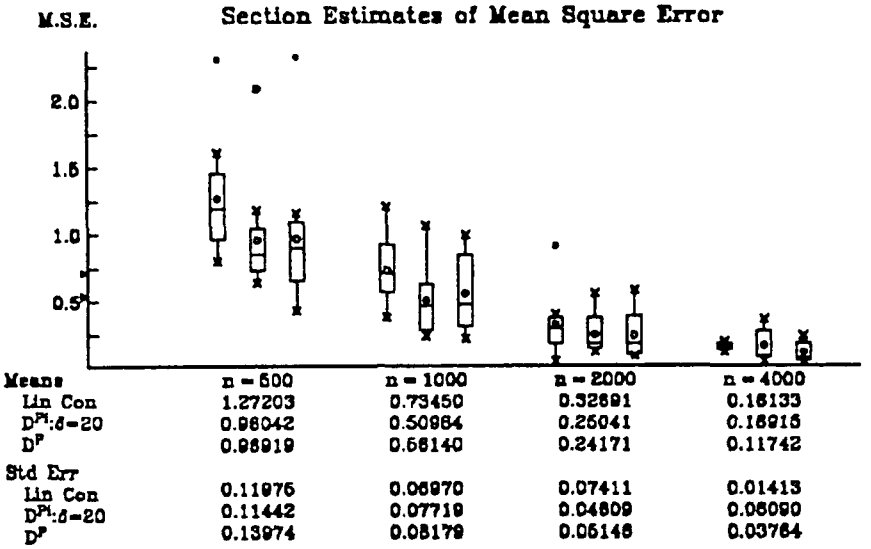
Control		Estimated Mean Square Error (% of Crude)			
	Estimator	n=500	n=1000	n = 2000	n=4000
	Crude	.06303	.08313	.09369	.10040
	Lin Con	.00664 (11)	.00611 (7)	.00565 (6)	.00734 (7)
1	$C + C^2$	.77137 (1223)	.59955 (715)	.43578 (465)	.38132 (379)
2	D^p	.00468 (7)	.00454 (5)	.00399 (4)	.00490 (5)
3	$D + D^p$	.01005 (7)	.00456 (5)	.00411 (4)	.00605 (6)
4	$D_i^{p_i}$	.00422 (7)	.00446 (5)	.00520 (6)	.01890 (13)

TABLE 4. Estimates of the mean square error of $\widehat{\overline{W'}}(m,n)$ for various n from 120,000 busy periods of an M/M/1 queue with traffic intensity .5. The number in () is 100 times the ratio of the controlled estimated mean square error to the crude estimated mean square error.

the mean square error and variance reduction, the 120,000 busy periods were separated into M independent "replications" so that multiple $\widehat{\overline{W'}}(m,n)$ could be calculated. For n = 500, the 120,000 busy periods separated into M = 12 replications; for n = 1000, 2000 and 4000, they were separated into 12, 10, and 5 replications respectively.

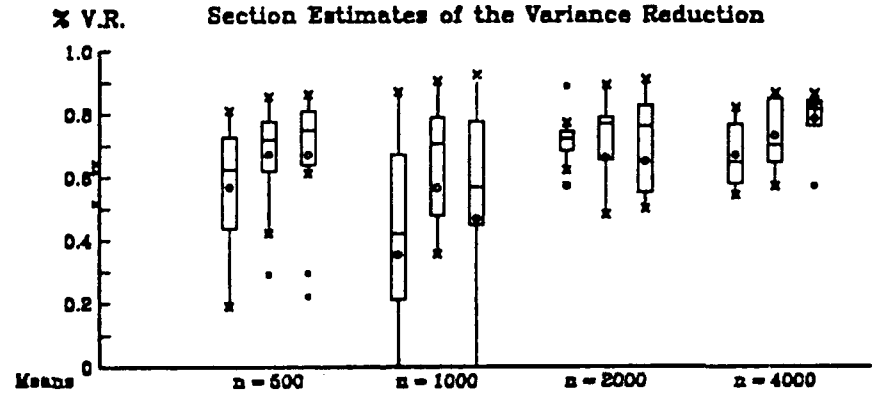
Figure 37 contains triples of boxplots of the estimated mean square error for each replication's $\widehat{\overline{W'}}(m,n)$ for the linear control and two of the nonlinear controls for different sample sizes n. Figure 38 contains triples of boxplots of the estimated variance reduction, computed as $1 - s^2_{\widehat{\overline{W'}}(m,n)} / s^2_{\widehat{\overline{W}}(m,n)}$, for each replication's $\widehat{\overline{W'}}(m,n)$ for the linear control and two of the nonlinear controls for different sample sizes n. For each triple of boxplots, the left one is the linear control, the middle is Control 4 and the right boxplot is Control 2. Figure 37 shows that as n increases to 4,000, the estimated mean square error decreases for each controlled estimate. It also shows that for each n, Control 2, D^p, tends to have a lower estimated MSE than the other controls.

Figure 38 shows that the nonlinear Control 2 is generally more effective at reducing the variance than the linear control. Additionally, the mean of the estimated variance reduction, .736 for n=4000, is close to .76, the largest estimated R^2 estimated by ACE. Unfortunately, as will be shown in the next section, when the traffic intensity increases to .99, the nonlinear control is no more effective than the linear control.



Means	n = 500	n = 1000	n = 2000	n = 4000
Lin Con	1.27203	0.73450	0.32691	0.16133
D^{PI}:δ=20	0.96042	0.50964	0.25041	0.16915
D^{P}	0.98919	0.56140	0.24171	0.11742
Std Err				
Lin Con	0.11976	0.06970	0.07411	0.01413
D^{PI}:δ=20	0.11442	0.07719	0.04809	0.06090
D^{P}	0.13974	0.08179	0.06146	0.03764

Figure 37. Boxplots of the estimated mean square error for multiple replications of three nonlinearly controlled estimates. For each triple of boxplots, the left one is the linear control, the middle is Control 4 and the right boxplot is Control 2.



Means	n = 500	n = 1000	n = 2000	n = 4000
Lin Con	0.56967	0.35502	0.57419	0.67108
D^{PI}:δ=20	0.87420	0.56697	0.68247	0.73244
D^{P}	0.87390	0.47007	0.65026	0.76867
Std Err				
Lin Con	0.06066	0.14620	0.16133	0.06108
D^{PI}:δ=20	0.04828	0.10350	0.09456	0.05523
D^{P}	0.06006	0.15450	0.10685	0.06397

Figure 38. Boxplots of the estimated variance reduction for multiple replications of three nonlinearly controlled estimates. For each triple of boxplots, the left one is the linear control, the middle is Control 4 and the right boxplot is Control 2.

2. The M/M/1 Queue with Traffic Intensity of .99

Another simulation experiment was conducted with the parameters chosen so that the traffic intensity would be .99 while the expected value of W remained at 10. The simulation was run until 200,000 busy periods were completed. In addition to the linear control, each of the four controls listed previously was evaluated for $\rho = .99$. The first part of the evaluation consisted of using the 200,000 busy periods to estimate $\overline{\widehat{W}'}(m,n)$, its variance and mean square error, for different n where $m \times n = 200{,}000$. None of the four controls did appreciably better at reducing the variance and means square error than the linear control of just C. Control 1 had problems with bias and Controls 2, 3, and 4 were ineffective at increasing the correlation of the transformed control with Z.

a. *The Performance of Control 1*

As one would suspect, Control 1 always had a higher estimated correlation with Z than did the linear control of just C. As an example, when the 200,000 busy periods were sectioned into 20 sections containing 10,000 busy periods each the estimate of the squared correlation between Z and the baseline linear control averaged .61 for the 20 sections; for Control 1, the average of the estimated square correlations was .89.

Unfortunately, the large increase in correlation for Control 1 did not translate into variance reduction or mean square error reduction. Table 5 shows the estimate of E[W], its estimated standard deviation and estimated mean square error as a function of n and the type of estimate for the crude, the linear control with C and Control 1. By comparing the rows for the standard deviation, one can see that while the Control 1 estimated standard deviation is less than the crude's estimated SD, it is generally larger than the baseline linear control's estimated SD.

The real difficulty with Control 1 can be seen by comparing the rows in Table 5 for $\overline{\widehat{W}}(m,n)$ with $\overline{\widehat{W}'}(m,n)$ and the associated rows for the estimate of the mean square error (MSE). The Control 1 estimate for $\overline{\widehat{W}'}(m,n)$ is clearly biased with its estimates of 10.4, 111.4, 11.8 and so on. As for the M/M/1 queue with traffic intensity .5, this bias causes the estimates of the mean square error to be greater for Control 1 than the crude estimate for n greater than 1000.

The source of the bias in the Control 1 estimates can be seen when the individual estimates $\widehat{W}'(n)_j$ are examined. For example, consider the 20 estimates $\widehat{W}'_j(10{,}000)$,

$j = 1, \ldots, 20$, that are used to compute the four estimates for $n = 10,000$ in the right-hand column for Control 1. These are the estimates described above in the first paragraph of this section that had an average estimate of squared correlation between Z and Control 1 of .89. However, 11 of the 20 estimates were further from the true value of 10 than their corresponding linear control C estimates and 8 of the 20 were further from 10 than their corresponding crude estimates. The most biased Control 1 estimates occurred when the crude estimate was at an extreme value. For crude estimates of 5.97, 7.11 and 17.26, the Control 1 estimates were 12.00, 17.31 and 12.18 respectively, despite having R^2 estimates from the regression that estimated the coefficients of .82, .94 and .95 respectively. These highly biased estimates for $\widehat{W}'_j(10,000)$ not only increase the bias of $\overline{\widehat{W}'}(m,n)$, they increase its estimated variance as well.

The high bias in the few individual $\widehat{W}'(n)$ can be attributed to bias in their control functions. While the estimate $\overline{C}$ was close to the true expected value of C for these sections, the estimate of $\overline{C^2}$ was not close to the true expected value despite a section size of 10,000. This can be attributed the higher variance associated with estimates of second moments such as C^2. The highly skewed nature of the distribution for C for a traffic intensity of .99 only serves to exacerbate the problem. In short, despite the fact that its estimated correlation with Z is much greater than for the linear control C, Control 1 is not as effective as using C as the linear control of the top because of the bias problem.

Table 5 also demonstrates that one can not rely solely on the reduction on the standard deviation as a measure of effectiveness of a control. The s'/s rows contain the ratio of the estimated standard deviation of the controlled estimate to that of the crude estimate. The row for Control 1 shows that Control 1 can be effective for this one sample at reducing the standard deviation through an appropriate choice of n. For $n = 4000$, Control 1 reduces the standard deviation of the crude estimate to 51% of its original value. However, at the same time, the bias of Control 1 causes a 383% increase in the estimate of the mean square error. Thus when evaluating controls for biased estimators, one must consider the effect of the control on the estimated mean square error in addition to its effect on the estimated standard deviation.

Crude

n	500	1000	2000	4000	5000	7000	8000	10,000
$\widehat{\overline{W}}(m,n)$	7.54	8.55	9.32	9.77	9.88	9.98	10.0	10.0
S.D.	.248	.336	.405	.485	.455	.526	.527	.550
MSE	6.10	2.50	.621	.287	.220	.277	.279	.304

The Baseline Linear Control (C)

$\widehat{\overline{W'}}(m,n)$	8.09	8.78	9.49	9.81	9.77	9.97	9.99	9.93
S.D.	.189	.232	.254	.308	.223	.268	.282	.228
MSE	3.67	1.54	.323	.133	.102	.072	.079	.057
s'/s	.76	.69	.63	.64	.49	.51	.54	.41

Control 1 ($C + C^2$)

$\widehat{\overline{W'}}(m,n)$	10.4	11.4	11.8	11.0	11.0	10.9	10.7	10.8
S.D.	.233	.256	.274	.248	.275	.318	.277	.402
MSE	.205	2.06	3.31	1.10	1.03	.878	.502	.762
s'/s	.94	.76	.68	.51	.60	.60	.53	.73

TABLE 5. Section estimates based on 200,000 busy periods for the stationary waiting time in an M/M/1 queue with traffic intensity of .99 for different sample sizes n.

b. *The Impact of the Choice of n*

Table 5 also shows the importance of selecting the proper number of busy periods n to use to calculate $\widehat{W}'(n)$. Iglehart and Lewis (1979) chose $n = 2000$ for their estimates. They noted that for $\rho = .99$ and $n = 2000$, the baseline linear control estimates $\widehat{W}'(n)_j$ were nonnormal and $\overline{\widehat{W}'}(m,n)$ had substantial bias. They recommended that n be increased beyond 2000 to alleviate these problems.

The rows for the baseline linear control of C in Table 5 indicate that using n greater than 2000 can reduce the bias of $\overline{\widehat{W}'}(m,n)$ as well as the estimate of its standard deviation. For each $n \geq 4000$, the estimates of $\overline{\widehat{W}'}(m,n)$, namely 9.81, 9.77, 9.97, 9.99, and 9.93, are clearly closer to the true value of 10 than the $n = 2000$ estimate of 9.49. Keeping in mind that, for a fixed total of busy periods, as the n of $\overline{\widehat{W}'}(m,n)$ increases the m necessarily decreases. Thus some of the estimated standard deviations of the higher n were larger than for $n = 2000$. However, the standard deviations were small enough so that the estimates of the mean square error range from 41% down to 17% of the $n = 2000$ estimated mean square error for the $n = 2000$ crude estimate.

One reason that the estimates of the standard deviation for the baseline linear controlled estimates did not increase with decreasing m is that the control became more effective at reducing the standard deviation. The ratio of controlled to crude estimates of the standard deviation, namely s'/s, for the baseline linear control at $n = 2000$ was .63; for $n = 5000$ it dropped to .49, and at $n = 10,000$ the s'/s ratio was .41.

Table 5 also shows the impact of the choice of n on a single sample of 200,000 busy periods. It indicates that choosing an n greater than 2000 when $\rho = .99$ can improve the overall effectiveness of the control at reducing the estimated mean square error of $\overline{\widehat{W}'}(m,n)$.

c. *The Performance of the Nonlinear Controls 2, 3 and 4*

Controls 2, 3 and 4 all depend upon transformations of D to increase their correlation with Z. For the M/M/1 queue with traffic intensity .5, Control 2 was able to increase its correlation with Z over the linear control of just C. For that simulation, with $\rho = .5$, the expected value of D was 2.22 as compared to an expected value of C of -15.5. Raising D to a power could have a definite effect on the shape of the transformed C. For $\rho = .99$ the expected value of D was only .025 as compared to an expected value for C

of -10.02. Raising D to a power or even using an exponential transformation of D had minimal effect on the shape of the transformed C. Thus for $\rho = .99$ the dominant factor in the control was the effect of τ, and D had little influence. This explains why Controls 2, 3 and 4, which involved transformations of D, were only as effective as the linear control of just C.

d. *Controlling the Bottom*

Since transformations of D were ineffective, that left only C and C^2 as possible controls. Controlling the top using Control 2, with C^2, had unavoidable bias problems. Another alternative is to control the bottom. Controlling the bottom is essentially the same as controlling the top as in (94). The difference is that the control function $H(C,\theta)$ is applied to τ on the bottom of the ratio instead of to Y on the top. It was hoped that by putting the control function on the bottom, its bias would not affect the ratio as badly as when it was used on the top. Unfortunately, the opposite effect occurred. Using Control 1 on the bottom yielded estimates that were more biased and more variable than the estimates produced by controlling the top. Thus controlling the bottom was not an improvement over controlling the top.

e. *Controlling the Ratio*

Another alternative to controlling the top is to control the ratio. Controlling the ratio involves using the control function to control the already formed ratio estimate. The thought behind controlling the ratio was that averaging the control function before applying it would reduce the bias of the control function, especially for Control 1, and thus reduce the bias of the controlled estimates.

Controlling the ratio is analogous to the straight-forward control of the mean in that the actual estimates $\widehat{W}(n)_j$ for $j,\ldots,m$ are controlled. Controlling the ratio is also analogous to control of a quantile estimate in that the estimates to be controlled as well as the control must be computed using data from n of the individual busy periods. Controlling the ratio uses the following expression to compute the controlled estimate:

$$\widehat{\overline{W'}}(m,n) = \frac{1}{m}\sum_{j=1}^{m}\left[\widehat{W}(n)_j - \frac{1}{n}\sum_{i=1}^{n} H\left(\underline{C}_i,\underline{\theta}\right)\right]. \tag{134}$$

Note that the control function is averaged prior to controlling the ratio. Thus one can use Control 1 with its known expected value for $\overline{(C^2)}$ and does not have to determine $\mathrm{E}\left[\left(\overline{C}\right)^2\right]$.

Three linear controls were evaluated for controlling the ratio. They were the baseline linear control of C, a control using just C^2 and Control 1 namely $C + C^2$. The most effective control in terms of both variance reduction and mean square error was the control of C^2.

In order to compare the ratio control with C^2 against the control of the top with C, the 200,000 busy periods were separated into eight independent "replications" of 25,000 each. Controlled estimates $\overline{\widehat{W}}_k(m,n)$ and $\overline{\widehat{W}'}_k(m,n)$ were calculated for various n for each of the $k = 1,\ldots,8$ replications. Associated with each point estimate of W was an estimate of the standard deviation and the mean square error of the point estimate.

Table 6 contains the averages of the eight estimated mean square errors (and their standard errors) along with the averages of the ratio of the standard deviation of the controlled estimate to the crude estimate (and their standard errors). The impact of the ratio control can be seen by comparing the averages of the s'/s ratios for the ratio control of C^2 to the averages for the top control by C. For each n the ratio control reduces the estimate of the standard deviation much more than the top control. However,by comparing the averages of the estimates of the mean square error, one can see that the average for the top control is generally smaller than that for the ratio control. The improvement in the variance reduction by the ratio control is offset by its bias problems. Thus the linear control of the top appears to create a better estimate when one considers both the variance reduction and the mean square error.

Unfortunately, for an M/M/1 queue with $\rho = .99$, the nonlinear transformations of D were no longer more effective than the linear control of the top. While the use of the transformation C^2 improved the reduction of the estimate of the standard deviation over using just C, the resulting bias problems, even when used as a ratio control, negated its effectiveness. Thus the most effective control for reducing the estimates of the standard deviation and the mean square error was the linear control of the top with C.

It appears that increasing n beyond 2,000 for the linear control, for a fixed total number of busy periods, reduced the estimated mean square error for the controlled estimate over using the smaller $n = 2,000$. While almost any reduction is useful for resource

Controlling the Top with C

n	2000	4000	5000	7000	8000
$\overline{\text{MSE}}$ (se)	1.1 (.22)	1.6 (.41)	.76 (.18)	1.18 (.43)	1.14 (.42)
$\overline{s'/s}$ (se)	.69 (.06)	.75 (.11)	.57 (.05)	.53 (.08)	1.01 (.33)

Controlling the Ratio with C^2 $(C + C^2)$

$\overline{\text{MSE}}$ (se)	1.51 (.56)	1.27 (.57)	1.54 (.68)	1.53(1.15)	1.53 (.68)
$\overline{s'/s}$ (se)	.49 (.07)	.49 (.07)	.45 (.08)	.40 (.11)	.58 (.09)

TABLE 6. Means and standard errors for estimates of the mean square error and the ratio of controlled to crude standard deviations for estimates of the stationary waiting time from 8 replications of 25,000 busy periods each of an M/M/1 queue with traffic intensity of .99.

intensive simulations such as a high traffic intensity queues, one would like to do better. The next section discusses a method for exploiting the special characteristics of the control of the top of the ratio estimator in order to drastically reduce the estimated mean square error of the estimate of the stationary waiting time.

D. AVERAGE REGRESSION-ADJUSTED CONTROLLED ESTIMATES FOR REGENERATIVE SIMULATIONS.

Average regression-adjusted controlled regenerative estimates result from using the regression-adjusted estimation technique of Heidelberger and Lewis (1981) in the context of controlled regenerative estimates. They developed the technique in the context of a generic, uncontrolled, regenerative simulation. The first subsection will present the technique in terms of the regenerative estimate of the stationary waiting time in a queue.

Section D.1, will briefly describe Heidelberger and Lewis's (1981 average regression-adjusted estimator. Section D.2, will discuss the incorporation of controlled estimates into the regression-adjusted technique and the impact using of ridge regression in lieu of least-squares regression. The techniques will be demonstrated using data from simulations of an M/M/1 queue and an M/G/1 queue. The data will show that by combining the techniques of regression-adjusted regenerative estimates and the linear control of regenerative estimates,

one can produce an estimate with substantially smaller variance and mean-square error than by either technique alone.

Section D.3 will discuss the impact of using independent average regenerative estimates in the regression-adjusted technique. The data from the M/G/1 queue simulation will be used as an example. It shows that the estimates produced by using fewer independent average regenerative estimates tend to have larger estimated mean square error than the estimates produced by using the standard technique with more, correlated, average regenerative estimates.

1. The Average Regression-adjusted Regenerative Estimate

Heidelberger and Lewis (1981) proposed the regression-adjusted technique in order to improve the analyst's ability to reduce the bias of a regenerative estimate (*re*) while assessing the normality/symmetry of the regenerative estimate. Their regression-adjusted technique exploits two aspects of the structure of regenerative simulations.

The first aspect of the structure is the i.i.d. nature of the busy periods. Since the busy periods are i.i.d., one can section a single simulation of $N = m \times n$ busy periods into m i.i.d. simulations of n busy periods each. Thus one can average the m estimates of $\mathrm{E}\left[\widehat{W}(n)\right]$, namely $\widehat{W}_j(n)$ for $j = 1, \ldots, m$, to get the *average regenerative estimate* (the *are*(m_k, n_k) of Heidelberger and Lewis 1981). The average *re* is nothing more than $\overline{\widehat{W}}(m,n)$ from (92). Heidelberger and Lewis's (1981) idea was to compute estimates $\overline{\widehat{W}}(m,n)$ for different values of n. Let n_k, for $k = 1, \ldots, p$, represent p different values of n. If for a given simulation of N busy periods one estimates $\overline{\widehat{W}}(m_k, n_k)$ for each of the p values of n_k, one gets p unbiased but correlated estimates of $\mathrm{E}\left[\widehat{W}(n_k)\right]$.

The second aspect that the regression-adjusted technique exploits is the known bias structure of the regenerative estimate i.e.,

$$\mathrm{E}\left[\widehat{W}(n)\right] = \beta_0 + \beta_1/n + \beta_2/n^2 + \cdots + \beta_d/n^d + \cdots. \tag{135}$$

Estimating the coefficients in (135) to eliminate some of the bias in the regenerative estimate leads one to the regression-adjusted regenerative estimate.

Let $\widehat{W}\mathrm{ra}(N)$ represent the *regression-adjusted regenerative estimate* of the stationary waiting time based on a simulation of N busy periods (the *rare*(N) of Heidelberger and

Note that the type of control is important in determining the effectiveness of the combination. If one uses a control function that is unbiased for all sample sizes, and not just asymptotically unbiased, then the average of the controlled *re*'s will tend to be almost identical to the average of the crude regenerative estimates. Thus the regression will yield virtually the same estimates for $\widehat{W}\text{ra}(N)$ and $\widehat{W}'\text{ra}(N)$ and their associated estimates of their mean square error and all will be for naught. However, if one uses a control function that is only asymptotically unbiased, the effects of the small sample bias may be such that the average *re*'s are different for the controlled *re*'s and the crude *re*'s. In this case, the regression will yield different estimates for $\widehat{W}\text{ra}(N)$ and $\widehat{W}'\text{ra}(N)$ and the estimated mean square error for the average regression-adjusted controlled estimate can be much lower than for the average regression-adjusted (crude) estimate. This is especially true when one is controlling the top of a ratio estimator.

A potential difficulty with the regression-adjusted technique is the tendency for the least-squares regression matrix columns, composed of k rows of $1, 1/n, 1/n^2, \ldots, 1/n^d$, to be collinear. The collinearity can increase the variance of the regression-adjusted regenerative estimates. Johnson and Lewis (1989) presented results demonstrating that using ridge regression in lieu of least squares regression can diminish the impact of the collinearity and produce estimates with lower estimated mean square error. Ridge regression developed from the realization that although least-squares estimators are the minimum variance among linear estimators, "they are not in general minimum-mean-square-error estimators in that class." (Kendall and Stuart, 1979, p.92) In the examples that follow, average *ridge* regression-adjusted estimates were computed using the ridge regression technique of Dempster, Schatzoff and Wermuth (1972).

a. *An Example: Estimating the Stationary Waiting Time in an M/M/1 Queue*

The 200,000 busy periods from the previous simulation of the M/M/1 queue with ρ=.99 was used to evaluate the performance of the regression-adjusted controlled regenerative estimate, $\overline{\widehat{W}'\text{ra}}(M, N)$, against both the section controlled estimate $\overline{\widehat{W}'}(m, n)$ and the average regression-adjusted crude estimate $\overline{\widehat{W}\text{ra}}(M, N)$. Other factors in the evaluation were the degree, $d = 1$ and $d = 2$, the type of regression, least-squares versus ridge

Lewis 1981). The estimate $\widehat{W}\text{ra}(N)$ is defined as the estimate of β_0 in (135). To estimate β_0, the p average regenerative estimates $\widehat{\overline{W}}(m_k, n_k)$ are used as dependent variables in an unweighted least-squares linear regression on $\beta_0 + \beta_1/n + \cdots + \beta_d/n^d$. The regression can be to order d=1, 2, or 3 or more. For a given order d, the regression-adjusted regenerative estimate $\widehat{W}\text{ra}(N)$ is unbiased out to terms of order $1/n^d$.

One needs an estimate of the variance of the regression-adjusted estimate though. Given that one can calculate a regression-adjusted regenerative estimate from a simulation of N busy periods, the final step of obtaining a variance estimate requires M independent replications of the regression-adjusted regenerative simulation. Thus in essence, one runs the simulation until a total of $M \times N$ busy periods are completed. Let $\overline{\widehat{W}\text{ra}}(M, N)$ denote the *averaged regression-adjusted regenerative estimate* formed from M replications of N busy periods each (the *arare(m, n)* of Heidelberger and Lewis, 1981). The estimate $\overline{\widehat{W}\text{ra}}(M, N)$ is simply the average of the M independent regression-adjusted estimates. Since $\overline{\widehat{W}\text{ra}}(M, N)$ is a sample mean, one can also estimate the variance of $\overline{\widehat{W}\text{ra}}(M, N)$ as the sample variance of the M regression-adjusted estimates divided by M.

An immediate concern with forming average regression-adjusted regenerative estimates is determining appropriate values for the various parameters such as M, N, p, the n_k and d. Heidelberger and Lewis (1981) describe a graphical protocol which can assist the analyst in selecting some of these values. For the remainder of this chapter, assume that the total number of busy periods in the simulation, namely $M \times N$, has been set at 200,000. The next subsection will discuss the methods for using the regression-adjusted technique with controlled regenerative estimates and the impact of ridge regression in lieu of least-squares regression.

2. Using the Regression-adjusted Technique with Controlled Estimates

Average regression-adjusted controlled regenerative estimates result from applying the regression-adjusted technique to controlled re's. The overall procedure is the same as described in the last section, Section D.1. However, instead of using $\widehat{W}(n)$ to calculate the average re, one uses $\widehat{W}'(n)$ to calculate the average controlled re. The notation for the average regression-adjusted controlled regenerative estimate is simply $\overline{\widehat{W}'\text{ra}}(M, N)$.

regression, and N, the number of busy periods used for computing each regression-adjusted controlled estimate.

For the remainder of this chapter the term "best" estimate will refer to the estimate which has the smallest estimated mean square error (MSE). Unfortunately, resource limitations precluded running multiple replications of 200,000 busy periods so that the variability of the estimates of the mean square error could be determined. Thus the judgement of "best" is based upon a single estimate of the mean square error. However, while the data will not be able to establish which particular parameters are optimal, it will establish trends that demonstrate the effectiveness of using the regression-adjusted technique in combination with controlled regenerative estimates.

Table 5 on Page 129 shows the best section crude estimate was $\overline{\widehat{W}}(40,5000)$ with an estimated MSE of .220. It also shows the best sectioned controlled estimate was $\overline{\widehat{W}}(20,10000)$ with an estimated MSE of .057.

Table 7 and Table 8 contain average regression-adjusted estimates of the stationary waiting time, crude and controlled respectively, along with estimates of their standard deviation (SD) and mean square error (MSE). Both tables indicate that for a fixed number of busy periods equal to $M \times N$, the choice of large M versus large N is important. In both tables, the estimates of the MSE in the row for $M = 8$ are each lower than the estimates in the rows for $M = 5$ and $M = 4$. This indicates that it is more important to have multiple regression-adjusted estimates (large M) than to have many regenerative estimates for forming the average regenerative estimates used in the regression (large N).

A second trend in the two tables is that for both the least squares and the ridge regression estimates, the degree $d = 1$ regressions produce better MSE estimates than the degree $d = 2$ regressions. For example, in the $M = 8$ row in Table 7, the least squares estimate of the MSE goes from .292 to .457 as d goes from 1 to 2 and the ridge regression estimated MSE in Table 7 goes from .265 to .267. These estimates also show that increasing the degree of regression from 1 to 2 caused a much larger increase in the estimated MSE for the least square regression estimate than for the ridge regression estimate.

Finally, in both tables the ridge regression at degree $d = 1$ produced the best average regression-adjusted estimate. For the average regression-adjusted (crude) estimate,

the $M = 8$ row in Table 7 had the best estimated MSE of .265. This was larger than the best section crude estimate from Table 5 of .220. However, the best average regression-adjusted linearly controlled estimate, the $M = 8$ row in Table 8, had an estimated MSE of .017. *This estimate is just 8% of the best sectioned crude estimate.* The average (least-squares) regression-adjusted estimate from the same row has an estimated MSE of .02, again less than 10% of the sectioned crude estimate.

In summary, for this simulation of the M/M/1 queue with traffic intensity of .99, combining the regression-adjusted technique with the technique of linearly controlled regenerative estimates produced dramatic decreases in the estimated mean square error for the estimates of the stationary waiting time. Next, the same techniques will be applied to estimates of the stationary waiting time from a simulation of an M/G/1 queue.

b. *An Example: Estimating the Stationary Waiting Time in an M/G/1 Queue*

As a second example of the use of the combination of the regression-adjusted technique with the controlled regenerative estimates, an M/G/1 queue was simulated for 200,000 busy periods. The same data, namely Y_i, τ_i, and D_i for $i = 1, \ldots, 200{,}000$ was collected in order to produce estimates of the stationary waiting time. The service times for the M/G/1 queue were distributed as independent variates from a Gamma distribution with shape parameter 1/2. The means of the Gamma distribution and the Exponential interarrival time distribution were selected so that known expected value of the stationary waiting time was again 10.0 and (due to resource constraints) the traffic intensity was .975.

The control used to compute the controlled regenerative estimates was of the same form as the baseline linear control used for the M/M/1 queue, namely C from (95). The expected value of C for the M/G/1 queue was computed using methods suggested by Iglehart and Lewis (1979). Average regression-adjusted controlled estimates for the stationary waiting time in an M/G/1 queue were calculated and compared against section crude estimates, section controlled estimates and average regression-adjusted (crude) regenerative estimates. Table 9 contains the estimates from the section crude and controlled estimates, namely $\overline{\widehat{W}}(m,n)$ and $\overline{\widehat{W}'}(m,n)$ calculated using (92) and (94), along with their associated estimated SD and MSE. 8 Comparing Table 5 with Table 9 one can detect several trends. The first is that the $\overline{\widehat{W}}(m,n)$ and $\overline{\widehat{W}'}(m,n)$ estimates for the M/G/1 queue in Table 9 tend

n= 500 1000 2000 4000 5000 7000 8000					
M, N		Least Squares		Ridge	
		$d=1$	$d=2$	$d=1$	$d=2$
8, 25,000	$\widehat{\overline{W}}\mathrm{ra}(M,N)$	10.2	10.4	10.1	10.2
	S.D.	.501	.509	.500	.490
	MSE	.292	.457	.265	.267
5, 40,000	$\widehat{\overline{W}}\mathrm{ra}(M,N)$	10.1	10.4	10.1	10.2
	S.D.	.593	.633	.589	.606
	MSE	.369	.532	.358	.415
4, 50,000	$\widehat{\overline{W}}\mathrm{ra}(M,N)$	10.1	10.4	10.1	10.3
	S.D.	.607	.645	.600	.620
	MSE	.387	.548	.379	.455

TABLE 7. Average regression-adjusted crude estimates based on 200,000 busy periods for the stationary waiting time in an M/M/1 queue with traffic intensity of .99.

n= 500 1000 2000 4000 5000 7000 8000					
M, N		Least Squares		Ridge	
		$d=1$	$d=2$	$d=1$	$d=2$
8, 25,000	$\widehat{\overline{W}'}\mathrm{ra}(M,N)$	10.0	10.2	10.0	10.1
	S.D.	.137	.149	.130	.140
	MSE	.020	.061	.017	.024
5, 40,000	$\widehat{\overline{W}'}\mathrm{ra}(M,N)$	10.03	10.	10.0	10.1
	S.D.	.180	.209	.176	.201
	MSE	.033	.073	.031	.054
4, 50,000	$\widehat{\overline{W}'}\mathrm{ra}(M,N)$	10.0	10.2	10.0	10.0
	S.D.	.175	.220	.175	.230
	MSE	.032	.080	.031	.065

TABLE 8. Average regression-adjusted linearly controlled estimates based on 200,000 busy periods for the stationary waiting time in an M/M/1 queue with traffic intensity of .99.

Crude

n	500	1000	2000	4000	5000	7000	8000	10,000
$\widehat{\overline{W}}(m,n)$	8.28	9.01	9.41	9.69	9.80	9.91	9.91	9.94
S.D.	.227	.280	.311	.359	.340	.353	.328	.288
MSE	3.02	1.07	.449	.244	.154	.134	.116	.083

Control by C

$\widehat{\overline{W}}'(m,n)$	8.69	9.23	9.54	9.74	9.85	9.87	9.94	10.06
S.D.	.164	.198	.200	.250	.245	.257	.251	.272
MSE	1.74	.631	.251	.136	.082	.082	.066	.078

TABLE 9. Section estimates based on 200,000 busy periods for the stationary waiting time in an M/G/1 queue with traffic intensity of .975.

to have more bias at the higher n than the estimates for the M/M/1 queue in Table 5. At the same time, the estimates of the standard deviation are generally smaller for the M/G/1 estimates than for the M/M/1 estimates.

The combination of the increased bias and smaller standard deviation of the M/G/1 queue estimates can be seen in the estimates of the MSE. The estimated MSEs for the M/G/1 crude estimates are always smaller than those for the M/M/1 queue. The best crude estimate in Table 9 occurs for $n = 10,000$ with an estimated MSE of .083, much less than the .220 at $n = 5000$ in Table 5. The best controlled estimate for the M/G/1 queue in Table 9 occurs at $n = 8,000$ with an MSE estimate of .066. This is greater than the best M/M/1 queue estimate of .057 at $n = 10,000$. Thus the linear control is less effective at reducing the MSE of the estimate of the stationary waiting time for this simulation of an M/G/1 queue at $\rho = .975$ than it is for the simulation of the M/M/1 queue at $\rho = .99$

Table 10 and Table 11 contain the estimates for the average regression-adjusted estimates both crude and controlled. The tables have the same layout as Tables 7 and 8 with regard to the choice of M and N and the type and degree of regression.

The results for the M/G/1 queue again show that the combination of the regression-adjusted technique and the controlled regenerative estimates is effective at reducing the estimate of the mean square error for the estimate of the

stationary waiting time. As with the M/M/1 queue estimates in Tables 7 and 8, the predominant factor for producing estimates with less MSE is the choice of M, the number of regression-adjusted estimates that are averaged, versus N, the number of busy periods used to produce each regression-adjusted estimate.

For the M/M/1 queue in Table 8, the estimated MSE clearly increased as M decreased. The row with $M = 8$ in Table 8 always had the best estimates. The tradeoff between M and N is not as clear cut for the M/G/1 queue. The smallest estimated MSE for the crude estimate in Table 10 is .070, which occurs for both $M = 8$ and $M = 5$. In fact three of the four estimates of the MSE are lower for the $M = 5$ row than for the $M = 8$ row. When M drops to $M = 4$, the MSE estimates usually double. Thus even though the larger M does not necessarily produce the best the average regression-adjusted (crude) estimate, one does not want M to be too small. For the average regression-adjusted controlled estimates in Table 11, the larger M, at $M = 8$, has the best MSE for each of the other factors, although not always by much. Again the row for $M = 4$ clearly has the highest estimated MSEs.

For the M/M/1 queue, increasing the degree of the regressions also increased the estimated MSE. For the M/G/1 queue, Tables 10 shows a similar but less drastic effect for the crude estimate. Increasing d to $d = 2$ tended to increase the estimated MSE. However, for the controlled estimates, Table 11 shows that increasing d to $d = 2$ improved the estimated MSE over $d = 1$.

For the M/M/1 queue, using ridge regression improved the estimated MSE. For the M/G/1 queue, using ridge regression had mixed effects. While sometimes it increased the MSE estimate over the least-squares MSE estimate, other times it decreased the MSE estimate. For no combination of factors did using ridge regression have a dramatic effect for the M/G/1 queue.

In summary, the lowest section crude estimate of the MSE was .083, for $n = 10{,}000$ in Table 9. The best section controlled estimate of the MSE was .066, for $n = 8000$ in the same table. The best average regression-adjusted (crude) estimate had an estimated MSE of .070, in two places in Table 10. While this is an improvement over the section crude estimate, it is not as good as the section controlled estimate. However, the best average regression-adjusted controlled estimate had an estimated MSE of .023.

This is a reduction of the estimate of the MSE to 33% of the MSE of the best crude estimate. While not as dramatic as the reduction to 10% achieved for the M/M/1 queue, combining the regression-adjusted technique with the controlled regenerative estimates still produced substantial reductions in the estimates of the mean square error of the estimates of the stationary waiting time.

The reductions achieved thus far were obtained by using the correlated average-regenerative estimates in the regressions. The next section will discuss modifying the regression-adjusted technique to use independent average regenerative estimates.

3. Using of Independent Average Regenerative Estimates

Heidelberger and Lewis's (1981) regression-adjusted technique described previously computes average regenerative estimates for different n using the same busy periods over and over. Assume N is 12,000 busy periods and the set of n_k for $k = 1,2,3$ is $\{500, 1000, 2000\}$. One would first use all 12,000 busy periods to compute the average regenerative estimate $\overline{\widehat{W}}(24,500)$. Then the same 12,000 busy periods would be used in a similar manner to compute $\overline{\widehat{W}}(12,1000)$ and $\overline{\widehat{W}}(6,2000)$. Since the average re for a given n, say $n = 500$, uses the same data as the average re's for the other values of n, 1,000 and 2,000, the set of average regenerative estimates used as the dependent variables in the regression are necessarily correlated with each other. The presence of pairwise correlation can mean that the least squares estimate of β_0, namely $\widehat{W}\text{ra}(N)$, is not the minimum variance estimator (Kendall and Stuart, 1979, p. 83).

One method for removing the pairwise correlation is to only use each busy period one time i.e., for computing a single n_k's average regenerative estimate. For the example of the 12,000 busy periods above, one would use the first 4,000 busy periods to compute $\overline{\widehat{W}}(8,500)$, the next 4,000 busy periods to compute $\overline{\widehat{W}}(4,1000)$, and the last 4,000 busy periods to compute $\overline{\widehat{W}}(2,2000)$. By using each busy period one time, the average regenerative estimates used as the dependent variables in the regression are independent.

For a given N and set of n_k however, the cost of independence is a decrease in the corresponding m_k, the number of regenerative estimates averaged to get $\overline{\widehat{W}}(m_k, n_k)$. Given the n_k are fixed, the variance of $\overline{\widehat{W}}(m_k, n_k)$ is a decreasing function of m_k; as m_k decreases, the variance of the $\overline{\widehat{W}}(m_k, n_k)$ increases. Thus the variance of the estimate

n= 500 1000 2000 4000 5000 7000 8000					
		Least Squares		Ridge	
M, N		$d=1$	$d=2$	$d=1$	$d=2$
8, 25,000	$\widehat{\overline{W}}\mathrm{ra}(M,N)$	9.94	10.0	9.89	9.88
	S.D.	.259	.264	.255	.252
	MSE	.071	.070	.077	.077
5, 40,000	$\widehat{\overline{W}}\mathrm{ra}(M,N)$	9.94	10.0	9.92	9.99
	S.D.	.258	.293	.258	.270
	MSE	.070	.087	.073	.073
4, 50,000	$\widehat{\overline{W}}\mathrm{ra}(M,N)$	9.97	10.1	9.97	10.1
	S.D.	.388	.403	.388	.399
	MSE	.151	.169	.157	.162

TABLE 10. Average regression-adjusted crude estimates based on 200,000 busy periods for the stationary waiting time in an M/G/1 queue with traffic intensity of .975.

n= 500 1000 2000 4000 5000 7000 8000					
		Least Squares		Ridge	
M, N		$d=1$	$d=2$	$d=1$	$d=2$
8, 25,000	$\widehat{\overline{W'}}\mathrm{ra}(M,N)$	9.94	10.0	10.0	9.91
	S.D.	.172	.150	.162	.150
	MSE	.033	.023	.038	.031
5, 40,000	$\widehat{\overline{W'}}\mathrm{ra}(M,N)$	9.95	10.0	9.92	9.94
	S.D.	.200	.182	.195	.174
	MSE	.042	.034	.044	.033
4, 50,000	$\widehat{\overline{W'}}\mathrm{ra}(M,N)$	9.96	10.0	9.95	9.99
	S.D.	.257	.219	.259	.220
	MSE	.067	.050	.069	.049

TABLE 11. Average regression-adjusted controlled estimates based on 200,000 busy periods for the stationary waiting time in an M/G/1 queue with traffic intensity of .975.

$\widehat{W}\text{ra}(N)$ depends upon the balance between reducing m_k, increasing the variance, and eliminating the correlation between the average re's, decreasing the variance.

To demonstrate the impact of using independent average regenerative estimates, the 200,00 busy periods from the simulation of the M/G/1 queue with traffic intensity .975 from the last subsection were used to calculate average regression adjusted estimates, both crude and controlled. Other factors included the type of regression, least-squares versus ridge, the degree of regression, $d = 1,\ldots,5$, and finally, the choice of the n_k. The set of values for n_k dictates N and thus M for the following reason. One would like the independent average regenerative estimates used in the regression to have approximately equal variance. Thus one should use the same number of busy periods to compute the average regenerative estimate for each n_k. For K values of n, one gets that $N = K \times n_K$ and $M = \lfloor 200,000/N \rfloor$.

The following four tables, Tables 12, 13, 14, and 15 contain the estimates of the stationary waiting time from the 200,000 busy periods. The layout of each table is the same. Each table covers a single type of estimate, crude or controlled, and a single type of regression, least-squares or ridge. Within each table, the same four sets of n_k and their associated M and N are used to calculate the estimates. The average regression-adjusted estimates and their associated estimates of the SD and MSE are displayed for each degree of regression.

It appears that a major factor in obtaining a low estimated MSE is using a large n_K. For each table, the best estimated MSE for degrees $d = 1, 2$, and 3 occurred for $M = 3$ where $n_K = 8,000$. For degree d=4 or 5 the best estimated MSE occurred for either $M = 3$ or $M = 6$, when $n_K = 4,000$. Unfortunately, the high n_K necessitates a low M and the low M implies that the estimates of the SD and MSE are highly variable. However, when n_K was reduced to 2,000, the MSE estimates usually doubled, despite M being as high as 14.

The impact of the degrees of regression is closely associated with the type of regression. The best least-squares estimates usually occurred when $d = 2$. The best ridge regression estimates usually occurred for $d = 3$. Using ridge regression did not lower the estimated MSE unless the degree of regression was 3 or higher. For degrees $d = 4$ and $d = 5$, ridge regression usually provided better estimated MSE than obtained using least-squares regression.

M=3,N=56,000 n = 125 250 500 1000 2000 4000 8000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{\overline{W}}\mathrm{ra}(M,N)$	9.69	10.1	10.3	11.3	10.8
S.D.	.364	.309	.192	.020	.824
MSE	.229	.105	.129	1.19	1.32

M=6,N=32,000 n = 40 80 125 250 500 1000 2000 4000

$\widehat{\overline{W}}\mathrm{ra}(M,N)$	8.99	9.81	10.1	10.5	10.9
S.D.	.279	.337	.400	.499	.898
MSE	1.10	.150	.170	.459	1.62

M=7,N=28,000 n = 80 125 250 500 1000 2000 4000

$\widehat{\overline{W}}\mathrm{ra}(M,N)$	9.33	9.72	9.82	9.75	8.22
S.D.	.341	.420	.498	.772	2.09
MSE	.565	.209	.280	.658	7.54

M=14,N=14,000 n = 40 80 125 250 500 1000 2000

$\widehat{\overline{W}}\mathrm{ra}(M,N)$	8.83	9.40	9.86	9.99	9.51
S.D.	.247	.400	.577	.972	2.19
MSE	1.43	.520	.350	.945	5.04

TABLE 12. Average independent least-squares regression-adjusted crude estimates based on 200,000 busy periods for the stationary waiting time in an M/G/1 queue with traffic intensity of .975.

The bottom line is that the average regression-adjusted controlled estimates of the stationary waiting time in Tables 14 and 15, which were produced using the independent average re's, have a higher estimated MSE than those in Table 11, which were produced using the correlated average re's. The best estimated MSE in Table 11 is .023, for $M = 8$ and $d = 2$, and the best estimated MSE for the estimates created using independent average re's is .033, in Table 15 for $M = 3$ and $d = 3$. Thus it appears that using fewer independent average regenerative estimates in the regression is not as effective at reducing the estimated mean square error as using more, correlated, average regenerative estimates.

M=3,N=56,000 n = 125 250 500 1000 2000 4000 8000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{W}\text{ra}(M,N)$	9.59	9.75	9.82	10.5	11.2
S.D.	.357	.248	.123	.116	.196
MSE	.296	.124	.048	.263	1.48

M=6,N=32,000 n = 40 80 125 250 500 1000 2000 4000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{W}\text{ra}(M,N)$	8.80	9.38	9.54	10.0	10.5
S.D.	.268	.316	.349	.347	.506
MSE	1.51	.484	.333	.120	.506

M=7,N=28,000 n = 80 125 250 500 1000 2000 4000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{W}\text{ra}(M,N)$	9.07	9.20	9.41	9.42	9.25
S.D.	.368	.435	.351	.498	1.19
MSE	1.00	.829	.471	.584	1.08

M=14,N=14,000 n = 40 80 125 250 500 1000 2000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{W}\text{ra}(M,N)$	8.54	8.73	9.51	9.93	10.3
S.D.	.220	.309	.443	.633	1.29
MSE	2.18	1.71	.436	.406	1.75

TABLE 13. Average independent ridge regression-adjusted crude estimates based on 200,000 busy periods for the stationary waiting time in an M/G/1 queue with traffic intensity of .975.

M=3,N=56,000 n = 125 250 500 1000 2000 4000 8000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{\widehat{W'}}\text{ra}(M,N)$	9.77	10.0	10.2	11.0	10.4
S.D.	.275	.215	.090	.020	.159
MSE	.129	.046	.048	1.00	.185

M=6,N=32,000 n = 40 80 125 250 500 1000 2000 4000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{\widehat{W'}}\text{ra}(M,N)$	9.12	9.69	9.70	9.87	9.91
S.D.	.173	.187	.168	.160	.465
MSE	.804	.135	.118	.043	.220

M=7,N=28,000 n = 80 125 250 500 1000 2000 4000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{\widehat{W'}}\text{ra}(M,N)$	9.48	9.79	9.88	9.89	9.62
S.D.	.234	.338	.415	.574	1.67
MSE	.325	.158	.187	.342	2.93

M=14,N=14,000 n = 40 80 125 250 500 1000 2000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{\widehat{W'}}\text{ra}(M,N)$	9.17	9.65	10.0	10.1	9.63
S.D.	.194	.300	.433	.639	1.37
MSE	.727	.213	.187	.418	2.01

TABLE 14. Average independent least-squares regression-adjusted controlled estimates based on 200,000 busy periods for the stationary waiting time in an M/G/1 queue with traffic intensity of .975.

M=3,N=56,000 n = 125 250 500 1000 2000 4000 8000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{\overline{W'\text{ra}}}(M,N)$	9.71	9.83	9.94	10.4	10.8
S.D.	.287	.203	.170	.091	.077
MSE	.166	.070	.033	.168	.646

M=6,N=32,000 n = 40 80 125 250 500 1000 2000 4000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{\overline{W'\text{ra}}}(M,N)$	9.01	9.39	9.39	9.69	9.84
S.D.	.178	.207	.188	.123	.191
MSE	1.01	.415	.407	.111	.062

M=7,N=28,000 n = 80 125 250 500 1000 2000 4000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{\overline{W'\text{ra}}}(M,N)$	9.32	9.44	9.59	9.83	9.47
S.D.	.249	.309	.292	.368	1.21
MSE	.524	.409	.253	.164	1.75

M=14,N=14,000 n = 40 80 125 250 500 1000 2000

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\widehat{\overline{W'\text{ra}}}(M,N)$	9.01	9.20	9.80	10.1	9.67
S.D.	.108	.255	.348	.432	.827
MSE	.992	.703	.161	.197	.792

TABLE 15. Average independent ridge regression-adjusted controlled estimates based on 200,000 busy periods for the stationary waiting time in an M/G/1 queue with traffic intensity of .975.

E. SUMMARY

This chapter has discussed various methods for using the regenerative method of simulation to estimate the stationary waiting time in a queue. The first method built upon Iglehart and Lewis's (1979) linear control for regenerative estimates by incorporating nonlinear transformations of the prospective controls. The expected value for several nonlinear controls for an M/M/1 queue were determined. These controls were then applied to data from simulations of M/M/1 queues with traffic intensities of .5 and .99. A nonlinear control was effective at reducing the standard deviation of the estimate of the stationary waiting time for the queue with the .5 traffic intensity. The same nonlinear control was ineffective at the .99 traffic intensity. While a control would improve the estimate of R^2 from the regression that estimates the coefficients for the control function, this R^2 was an unreliable predictor of the effectiveness of the control, especially since it neglects the impact of the bias of the control on the bias of the estimate.

The bias of the linearly and nonlinearly controlled estimates necessitated using the estimated mean square error as a measure of effectiveness. Otherwise a technique that provided substantial variance reduction at the cost of substantial bias would have been deemed effective. The estimate of the MSE was used to assess the effectiveness of combining the technique of linearly controlled regenerative estimates with Heidelberger and Lewis's (1981) regression-adjusted technique for regenerative estimates.

Combining the two techniques created an average regression-adjusted controlled estimate which had a much smaller estimated mean square error than that produced using either technique individually. This was demonstrated for simulations of both an M/M/1 and M/G/1 queue. The reduction of the MSE estimate to 10% or even 33% of the best crude estimate was a dramatic reduction for such resource intensive simulations as an M/M/1 queue with a traffic intensity of .99 and M/G/1 queue with a traffic intensity of .975.

Finally a modification to the regression-adjusted technique was discussed. It was demonstrated that for a fixed overall simulation size, it was better to use "many" correlated average regenerative estimates for the regressions than to use a "few" independent ones. The increase in the variance of the average regenerative estimates caused by decreasing the number of regenerative estimates used for the averaging overwhelmed

the potential reduction gained by satisfying one of the requirements to have a minimum variance estimator.

An additional factor in favor of the regression-adjusted technique is that it has an associated graphical analysis. This is given in Heidelberger and Lewis (1981) and is not discussed here.

VI. THESIS SUMMARY

This dissertation has presented several ideas oriented to improving methods for analyzing the statistical output from stochastic computer simulations. The central theme of this dissertation has been to provide a means for assessing the variability of an estimate from a simulation and to reducing the variance of the estimate i.e., increasing the precision of the estimate, without a major increase in the resources required for designing, running and analyzing the simulation. With these methods, one can either increase the precision of an estimate for a fixed amount of resources or reduce the expenditure of resources required to achieve a given precision.

The two major areas presented were the use of nonlinear controls for variance reduction and the use of average regression-adjusted controlled regenerative estimators. Nonlinear controls were discussed in each chapter and the regenerative estimators were discussed in Chapter V. It was shown that for some simulations, nonlinear controls could reduce the variance of the estimate as compared to using linear controls. For the more complex case of regenerative simulations which involve ratio estimators, where the nonlinear controls had difficulty, the average regression-adjusted controlled regenerative estimators had much smaller estimates of the mean square error than either regression-adjusted estimators or controlled estimators.

After the introduction in Chapter I, Chapter II discussed linear controls for variance reduction and showed that nonlinear transformations could improve their effectiveness. Nonlinear controls were then defined in the context of nonlinear transformations of a control variable. It was shown how Breiman and Friedman's (1985) ACE algorithm could be used to estimate the maximum correlation between a statistic of interest and a potential control. ACE provided a means for evaluating a control and suggesting possible nonlinear transformations for improving the effectiveness of the control. Various methods for introducing nonlinearity and methods for determining the resulting nonlinear control's expected value were presented. It was also shown that nonlinear least-squares regression could be used to estimate the optimal parameters for some of the nonlinear transformations. As an ex-

ample of the use of nonlinear controls for variance reduction, nonlinear controls were used to improve the precision of estimates of the mean of W_2^2, Anderson and Darling's (1952) goodness-of-fit statistic. The nonlinear controls were able to improve the precision over the linear controls and come close to the ACE estimate of the maximum achievable variance reduction.

Chapter III discussed the use of nonlinear controls for reducing the variance of quantile estimators. Order-statistic-based quantile estimators are a discontinuous function of the data. It discussed how the discontinuous nature causes some subtle differences between controlling the variance of quantile estimators and controlling estimates of a mean. It was shown how when the control is a quantile estimator, one can use its asymptotic expected value in the control function instead of its true expected value. It was demonstrated how the use of strictly monotonic transformations of a quantile estimator control can greatly reduce the difficulty of calculating the expected value of the transformed control.

It was demonstrated in Chapter III that the method of sectioning yields a more reliable estimate of the variance of the quantile estimate than the method of jackknifing. ChapterIII also discussed the asymptotic normality of quantile estimates as it pertains to selecting section sizes. For a fixed sample size, as the section size used to compute the quantile estimate increases, the quantile estimates become more normal and the linear control becomes optimal. However it is demonstrated that by using a nonlinear control at smaller section sizes, one can get a more precise quantile estimate than by using the linear control at the larger section size.

Chapter IV used asymptotic expansions for the moments of an order-statistic-based quantile estimator to develop additional asymptotic expansions. These expansions included the variance of a strictly monotone transformation of a quantile estimator and the covariance between two different strictly monotone transformations of the same quantile estimator. Using these new expansions, expansions were developed for the squared correlation between two strictly monotone transformations of a quantile estimator and the ratio of squared correlations between the two transformations of the quantile estimator. These expansions were used to prove that one can select a nonlinear strictly monotone transformation for a quantile estimator control such that the squared correlation between the nonlinearly transformed quantile and the quantile of interest is greater than the squared correlation

between the quantile interest and a linear transformation of the quantile control. Thus the nonlinear control is more effective at reducing the variance of the estimate of the quantile of interest than the linear control. The expansions were compared to estimates from an example simulation and found to be fairly accurate at predicting the variances, the squared correlations and the ratio of squared correlations.

Chapter V discussed the use of nonlinear controls in regenerative simulations of queueing systems. The chapter built on Iglehart and Lewis's (1979) linear control for regenerative estimates and Heidelberger and Lewis's (1981) regression-adjusted technique for regenerative estimates. Nonlinear controls were proposed for controlling the estimate of the stationary waiting time in an M/M/1 queue. Formulas for the expected values of nonlinearly transformed controls were developed. These nonlinear controls were tested using simulations of an M/M/1 queue with traffic intensities of .5 and .99. One of the nonlinear controls was an improvement over the linear control for the .5 traffic intensity estimate. Its variance reduction approached the ACE estimate for maximum achievable variance reduction. Unfortunately, when the traffic intensity increased to .99, none of the nonlinear controls were more effective than the linear control. A second linear control, which used the transformation C^2, greatly increased the estimate of R^2 given by the least-squares regression that estimated the coefficients for the control function. Unfortunately this R^2 estimate was a poor predictor of the reduction in variance of the controlled ratio estimate. It was also shown that one must use the estimate of the mean square error instead of the variance to judge the effectiveness of the controls. Some controls reduced the variance but at the cost of greatly increasing the bias of the estimate.

ChapterV also proposed using average regression-adjusted controlled regenerative estimates as an alternative to using nonlinear controls to reduce the estimated mean square error of the regenerative estimate of the stationary waiting time. These estimates result from using the controlled regenerative estimates to form the average regenerative estimates in the regression-adjusted technique of Heidelberger and Lewis (1981). Their technique is essentially a generalized jackknife which utilizes graphical analysis. These estimates were computed for the simulation of the M/M/1 queue with .99 traffic intensity as well as for a simulation of an M/G/1 queue with traffic intensity of .975.

The average regression-adjusted controlled regenerative estimates were computed using both least-square regression and ridge regression for several degrees of regression. It was shown that the choice of parameters can have a large effect on the estimated mean square error. It was also shown that the average regression-adjusted controlled estimate can have an estimated mean square error as low as 10% for the M/M/1 queue and 33% for the M/G/1 queue of the best estimated mean square error for a crude estimate. This was a dramatic reduction in the estimated mean square error, especially as it pertained to queueing resource intensive simulations.

In summary, this dissertation provides several methods for improving one's analysis of the output from a computer simulation. First, it shows how one can use the ACE algorithm to evaluate the effectiveness of a proposed control and suggest nonlinear transformations for improving the effectiveness of the control. It develops several methods for introducing nonlinearity into a control function and establishes that one can use nonlinear least-square regression for estimating the parameters of the transformation. It shows that nonlinear controls can be more effective than linear controls at reducing the variance of estimates of the mean as well as quantile estimates. Finally it demonstrates that while nonlinear controls may have limited effectiveness in regenerative queueing simulations, the use of linear controlled estimates to produce average regression-adjusted controlled regenerative estimates can dramatically reduce the estimated mean square error of the estimate of the stationary waiting time.

Nonlinear controls will not be applicable for many simulations. One must be able to compute the expected value of the transformed control to use them at all. However they do provide an option that one can investigate for analyzing the output of a simulation. They can improve the effectiveness of a control so as to make controlling an estimate a worthwhile procedure. The reductions achieved by the average regression-adjusted controlled regenerative estimates should make them a viable option for analyzing the output from regenerative simulations. These methods can work and they can improve the analysis of output from stochastic computer simulations.

LIST OF REFERENCES

Anderson, T. W. and Darling, D. A., "Asymptotic Theory of Certain "Goodness-of-fit" Criteria Based on Stochastic Processes", *Annals of Mathematical Statistics*, v. 23, pp. 193-212, 1952.

Bard, Y., *Nonlinear Parameter Estimation*, Academic Press, 1974.

Barndorff-Nielson, O. and Cox, D., *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, 1989.

Beale, E., "Regression: A Bridge Between Analysis and Simulation", *The Statistician*, v. 34, pp. 141-154, 1985.

Breiman, L. and Friedman, J. H., "Estimating Optimal Transformations for Multiple Regression and Correlation", *Journal of the American Statistical Association*, v. 80, no. 391, pp. 580-619, September 1985.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A., *Graphical Methods for Data Analysis*, Wadsworth & Brooks/Cole, 1983.

Cramér, H., *Mathematical Methods of Statistics*, Princeton University Press, 1966.

Crane, M. A. and Iglehart, D. L., "Simulating Stable Stochastic Systems: III. Regenerative Processes and Discrete-Event Simulation", *Operations Research*, v. 23, no. 1, pp. 33-45, January-February 1975.

David, H. A., *Order Statistics*, John Wiley & Sons, Inc., 1970.

David, F. N. and Johnson, N. L., "Statistical Treatment of Censored Data", *Biometrika*, v. 38, pp. 463-468, 1956.

Dempster, A. P., Schatzoff, M., and Wermuth, N., "A Simulation Study of Alternative to Ordinary Least Squares", *Journal of the American Statistical Association*, v. 72, pp. 77-106, 1972.

Efron, B. and Gong, G., "A Leisurely Look at the Bootstrap, the Jackknife and Cross-validation", *The American Statistician*, v. 37, no. 1, pp. 36–48, February 1983.

Gallant, A. R., *Nonlinear Regression*, John Wiley & Sons, 1987.

Glynn, P. W. and Whitt, W., "Indirect Estimation via $L = \lambda W$", *Operations Research*, v. 37, no. 1, pp. 82–103, 1989.

Heidelberger, P. and Lewis, P. A. W., "Regression-adjusted Estimates for Regenerative Simulations, with Graphics", *Communications of the ACM*, v. 24, no. 4, pp. 260–273, April 1981.

Heyman, D. P. and Sobel, M. J., *Stochastic Models in Operations Research*, v. I, McGraw Hill, 1982.

Hsu, J. C. and Nelson, B. L., "Control Variates for Quantile Estimation", In *Proceedings of the 1987 Winter Simulation Conference*, Thesen, A., Grant, H., and Kelton, W. D., editors, pp. 342–346, 1987.

Iglehart, D. L. and Lewis, P. A. W., "Regenerative Simulation with Internal Controls", *Journal of the Association for Computing Machinery*, v. 26, no. 2, pp. 271–282, April 1979.

Johnson, M. E., *Multivariate Statistical Simulation*, Wiley, N.Y., 1987.

Johnson, L. D. and Lewis, P. A., *Simulation and the Delta Method: Viable Estimation in Asymptotic Expansions*, The 21st Symposium on the Interface, 1989.

Kendall, M. and Stuart, A., *The Advanced Theory of Statistics*, 4th ed., v. 1, Macmillan, 1977.

Kendall, M. and Stuart, A., *The Advanced Theory of Statistics*, 4th ed., v. 2, Griffin, 1979.

Kleijnen, J. P. C., *Statistical Techniques in Simulation: Part I*, Dekker, 1974.

Kleinrock, L., *Queueing Systems Volume I: Theory*, v. 1, John Wiley & Sons, 1975.

Lancaster, H. O., "Kolmogorov's Remark on the Hotelling Canonical Correlation", *Biometrika*, v. 53, no. 3 and 4, pp. 585–590, 1966.

Lavenberg, S. S., Moeller, T. L., and Welch, P. D., "Statistical Results on Control Variables with Application to Queueing Network Simulation", *Operations Research*, v. 30, no. 1, pp. 182–202, January and February 1982.

Lavenberg, S. S. and Welch, P. D., "A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations", *Management Science*, v. 27, no. 3, pp. 322–335, March 1981.

Lewis, P. A. W. and Orav, J., *Simulation Methodology for Statisticians, Operations Analysts, and Engineers*, v. 1, Wadsworth & Brooks/Cole, 1989.

Marquadt, D. W., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters", *SIAM Journal*, v. 11, pp. 431–441, 1963.

Miller, R. G., "The Jackknife - A Review", *Biometrika*, v. 61, no. 1, pp. 1–15, 1974.

Mood, A. M., Graybill, F. A., and Boes, D. C., *Introduction to the Theory of Statistics*, 3rd ed., McGraw Hill, 1974.

Nelson, B. L., *Control-Variate Remedies*, Department of Industrial and Systems Engineering, The Ohio State University, Report 1988-004, October 1988.

Royden, H. L., *Real Analysis*, 3rd ed., Macmillan Publishing Company, 1988.

Rubinstein, R. Y. and Marcus, R., "Efficiency of Multivariate Control Variates in Monte Carlo Simulation", *Operations Research*, v. 33, no. 3, pp. 661–667, May-June 1985.

Sampson, A. R., "A Tale of Two Regressions", *Journal of the American Statistical Association*, v. 69, no. 347, pp. 682–689, 1974.

Seila, A. F., "A Batching Approach to Quantile Estimation in Regenerative Simulation", *Management Science*, v. 28, pp. 573–581, 1982.

Serfling, R. J., *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.

Smith, W. L., "Renewal Theory and Its Ramifications", *Journal of The Royal Statistical Society: Series B*, v. 20, pp. 243–302, 1958.

Tocher, K. D., *The Art of Simulation*, The English Universities Press, Ltd, 1963.

Weiss, L., "On the Asymptotic Joint Normality of Quantiles from a Multivariate Distribution", *Journal of Research of the National Bureau of Standards*, v. 68B, no. 2, pp. 65–66, April-June 1964.